3

WLAN QoS

Mathilde Benveniste[a]

With both the enterprise and residential sectors embracing voice over IP (VoIP) at an accelerating pace, and the pervasive use of wireless local area networks (WLANs), the natural requirement emerged for a technology to support VoIP over WLANs without degradation of its quality of service (QoS). QoS requirements for WLANs are imposed also by video and multimedia applications tailored for use with WLANs. A QoS-focused MAC Layer standard, IEEE 802.11e, was developed to meet the QoS requirements of a range of applications. In addition to VoIP/ multimedia QoS, the new standard serves mission-critical functions by reducing latency across a WLAN. This chapter discusses the enhancements the new standard adds to WLAN technology with respect to QoS performance, channel use efficiency, and power management of battery-based wireless devices.

3.1 Introduction

Since the initial emergence of the 802.11 network interface card for laptop computers and access points, the appeal of 802.11 technology has been so strong worldwide that it is now appearing in a wide range of devices, including consumer electronics devices and VoIP phones. Enterprises wish to extend VoIP over wireless LANs for the convenience wireless service brings to the mobile user throughout the building, campus, quad and warehouse, as well as anywhere a WLAN is accessible. Residential users purchasing VoIP service for cost savings, look to the WLAN to enable them to make their telephones cordless. The installation of WLANs in public spaces, backed up by a ubiquitous Internet, makes the case of VoIP over WLANs even more compelling. Users can have telephone service portability free of any wires anywhere a WLAN is present. The new trends in the expansion of WLAN use include consumer electronics appliances generating multi-media traffic streams from applications such as video streaming and interactive gaming.

"All this could happen if wireless LANs could support QoS adequately in a congested WLAN" was typically the reaction to the above observations prior to the adoption of the new standard for IEEE 802.11 WLANs, known as IEEE 802.11e. The new standard enables frames from QoS-sensitive applications to be transmitted sooner than other frames, thus minimizing latency. It also introduces new power management features that will prolong the life of mobile devices powered by battery. The channel-use efficiency gains

[a] *Avaya Labs*

.

introduced by the new standard make it worth pursuing even in situations where all traffic is of the same type, thus allowing privileged treatment to none. The lower latency achievable with 802.11e enables also the prioritization of time-critical data. Devices observing the new standard can co-exist with 802.11-compliant devices.

This chapter gives a high level overview of the major 802.11 mechanisms that have been modified and the new mechanisms introduced in 802.11e. They cover specifically the areas of channel access, admission control, and power management. QoS challenges that remain specifically in 802.11 mesh networks are also discussed.

3.1.1 Terminology and Abbreviations

For the reader reviewing the IEEE 802.11e standard [1], we note in this section some relevant naming conventions used in the standard. The QoS-aware contention-based random access is referred to as Enhanced Distributed Channel Access (EDCA). In the early standard drafts, and in much of the published literature, the same access approach had been called 'enhanced distributed coordination function' (EDCF), following the naming convention of the 802.11 standard, where Distributed Coordination Function (DCF) referred to contention-based random access [2]. Polled access was called Point Coordination Function (PCF) in the 802.11 standard. The enhanced polled access mechanism in 802.11e is called HCF Controlled Channel Access (HCCA). An access point (AP) supporting 802.11e features is called a 'QAP', and a station equipped to use 802.11e features is called a 'non-AP QSTA'. A QoS-aware WLAN, i.e., the group of stations supported by a QoS-aware AP, is called 'QBSS' in the 802.11e standard as compared to 'BSS' (basic service set), the group of stations supported by a legacy AP.

Prioritization for the various functions of the channel access protocol is achieved by imposing waiting requirements of variable durations after the channel becomes available. The different durations are known as interframe spaces, with the shortest, SIFS (short inter-frame space), used when a transmission is acknowledged, PIFS (priority inter-frame space) used for PCF, and DIFS (DCF inter-frame space) required of stations using DCF.

Abbreviations or acronyms used in this chapter are defined below.

AP	access point
AIFS	arbitration inter-frame space
APSD	automatic power save delivery
CSMA/CA	carrier sense multiple access/collision avoidance
CW	contention window
CWMax	contention window maximum
CWMin	contention window minimum
DCF	distributed coordination function
DIFS	DCF inter-frame space
EDCA	enhanced distributed channel access
HCCA	HCF controlled channel access
PIFS	priority inter-frame space
PCF	point coordination function
S-APSD	scheduled APSD
SIFS	short inter-frame space

TCMA tiered-contention multiple access
TSPEC traffic specification
TXOP transmit opportunity
U-APSD unscheduled APSD
WLAN wireless local area network

3.2 Channel Access

A Wireless LAN operates on either the 2.4 GHz ISM band or the 5 GHz UNII band, each containing multiple radio channels. The IEEE 802.11 standard specifies procedures for WLAN stations by which they share a single radio channel for asynchronous data transfer. Two channel access mechanisms are specified, contention-based and polled access. With contention-based access, stations transmit to peers and to the AP by accessing the channel using the distributed random access method that employs the CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance) MAC protocol [3]. If an AP is present in a WLAN, peer-to-peer communication is not allowed independently of the AP. With polled access, the AP transmits frames to a station and polls for its transmissions.

The IEEE 802.11e standard amendment also provides contention-based and polled access mechanisms, both of which represent enhancements of the 802.11 mechanisms. The latter are referred to in this chapter as the ‘legacy’ mechanisms. 802.11e aims at reducing access delay and jitter in delivering QoS-sensitive frames from the source to the destination through enhanced functionality at the MAC Layer. The access delay comprises over-the-air time and queuing delay plus time consumed in retransmissions, when they occur. To achieve this goal, channel access in an 802.11e-compliant WLAN distinguishes among priorities of individual frames as introduced by IEEE 802.1D [4]. QoS-sensitive traffic is typically assigned higher priority than best effort data. Stations may transmit/ receive traffic streams of different priorities concurrently. The channel access mechanisms are described in detail later in this section.

The 802.11e amendment also introduces features to improve channel use efficiency. It allows stations in WLAN served by an AP to communicate directly with one another. Signaling must be exchanged between the stations, through the AP, according to the Direct Link Setup protocol.
A block acknowledgement mechanism is introduced in 802.11e, which improves channel efficiency by aggregating several acknowledgements into a single frame. Special signaling is needed between stations to negotiate this type of acknowledgement.

Another efficiency enhancing feature introduced in 802.11e is the transmit opportunity (TXOP). In a contention based TXOP, a station may transmit a sequence of frames without having to contend for the channel, following a successful channel access attempt. Because all but one frame in a TXOP is transmitted without contention, TXOPs help reduce the frequency of collisions and thus increase channel use efficiency. A station granted a ‘polled TXOP’ when polled by the AP may transmit several frames to the AP, thus obviating the need for additional polls.

The remainder of this section describes the 802.11e channel access mechanism after first presenting the channel access protocols employed in 802.11, namely contention-based and polled access.

3.2.1 Legacy Channel Access Methods

3.2.1.1 Legacy Contention-Based Channel Access

According to the legacy contention-based access mechanism, DCF, each station listens to the channel and, if busy, postpones transmission and enters into the 'backoff procedure' [2]. This involves deferring transmission by a random time, which facilitates collision avoidance between multiple stations that would otherwise attempt to transmit immediately after completion of the current transmission.

The length of time for which a station will postpone its transmission depends on the 'backoff value', a number chosen randomly from a range of integers known as the Contention Window (CW). The backoff value expresses, in time slots, the cumulative time the channel must be idle before access may be attempted, excluding an additional time interval DIFS that the channel must remain idle following each period the channel is busy. In other words, transmission occurs after initially setting a 'backoff timer' to the backoff value, and then counting down once for every slot of time that the channel remains idle following a busy period excluding an initial idle period of length DIFS. A transmission may not be attempted until the backoff timer expires. CW is set initially at the value CWmin, and is doubled after each collision involving its transmitted frame, until reaching the value CWmax; after which it remains constant for further retries. The frame is dropped if it cannot be transmitted successfully after a specified number of retries. CW is reset to CWmin following a successful transmission. When the backoff timer expires, it is reset to a new backoff value, drawn randomly from the contention window CW, regardless of whether there is a frame queued for transmission.

Following anyone's transmission on the channel, a station is allowed to transmit only after the channel remains idle for at least a DIFS time interval. The value of DIFS is selected in order to enable DCF to share the same channel with the centralized protocol PCF of the 802.11 standard. In PCF, the channel is accessed at the end of a transmission after a PIFS idle interval, which is shorter than DIFS.

3.2.1.2 Legacy Polled Channel Access

At pre-specified regular time intervals, an AP engaged in polled access starts a 'contention free period' by transmitting a beacon frame. The AP can access and maintain control of the channel, once the channel becomes idle for the duration of the contention-free period by transmitting after a shorter time interval, PIFS, than a station. In addition, a station hearing the beacon will refrain from transmitting if not polled until it receives notice from the AP that the contention-free period is over.

The AP first transmits all broadcast and multicast frames and frames addressed to power-saving stations that subscribe to polled access. The transfer of frames to and from non-power-saving stations follows. The AP maintains a 'polling list' of stations to be polled, and polls each of them at least once during a contention-free period. At the same time the AP transfers frames to the stations on the polling list. Every poll elicits a single frame from the polled station. The AP stops polling a station in a contention-free period if it receives a Null frame (frame containing a header but no body) in response to a poll. An acknowledgement, data and/or poll can be combined into a single frame in order to save overhead.

3.2.2 802.11e Contention-Based Channel Access

The IEEE 802.11e contention-based access mechanism, EDCA, extends the contention-based access mechanism of the 802.11 standard to provide frame prioritization [1]. That is, given a collection of contending entities, prioritized access enables frames of higher priority to access the channel sooner. IEEE 802.11e uses the TCMA MAC protocol, a variant of CSMA/CA designed for priority differentiation [5, 6].

A key consideration in formulating EDCA was fairness. Because certain stations will transmit frames of different priorities, while others will transmit frames of a single priority, it was important for the channel access mechanism to provide the same performance to all frames of a given priority regardless of their source. Thus, instead of buffering all frames in a single queue (as with 802.11 stations), an 802.11e station employs four queues, one for each 'access category' based on the frame's priority [7]. The mapping of user priorities to access categories specified for a WLAN must be observed by all the stations. Traffic in an access category mapped to higher priorities will access the channel more readily than lower-priority access categories.

The different queues of a station contend for channel access independently of one another, almost as if they resided in different stations. The only difference is that any internal collisions between two queues of the same station are resolved by allowing the higher priority frame to be transmitted, while the lower priority frame is treated as if it had experienced a collision.

A contending queue in a station with multiple access categories behaves just like a station with traffic of a single access category with respect to accessing the channel. For simplicity of presentation, therefore, both are referred to as 'a station' in the description of prioritized access that follows.

The underlying MAC protocol in 802.11e contention-based access is CSMA/CA, which was described above. In 802.11e, the protocol's parameters CWmin and CWmax values are allowed to vary with the access category [8]. Assigning lower CWmin or CWmax values to an access category causes the contention window to be shorter when transmitting, or retransmitting following a collision. This indeed could offer priority differentiation, as shorter backoff values would lead to shorter average access delays, provided that the number of contending stations in the access category in question is small. The user should be cautioned, however, that when the number of contending stations in the access category is large, a short contention window would cause multiple stations to draw the same backoff, leading to collisions and consequently longer rather than shorter delays [9]. When differentiation with respect to CWmin or CWmax is pursued, the AP must be equipped to adjust these parameters dynamically in response to traffic conditions. The 802.11e standard permits such adjustments.

Another priority differentiation mechanism for contention-based access is through differentiation of the arbitration time, referred to as AIFS in the 802.11e standard. This concept was introduced as part of the TCMA MAC protocol, which is described in the next section [5, 6].

Like the contending stations, the AP also differentiates between traffic of different priorities by using different access parameters. The access parameters used by the AP for a given access category may be different from those used by the stations. The AP is thus allowed to use higher-priority access parameters than the stations, a prudent measure since the AP typically transmits more traffic than a station.

An important-efficiency enhancing feature introduced in 802.11e is the transmit opportunity [10]. Following a successful channel access attempt, a station may transmit a sequence of frames without having to contend for the channel. That is, a station is allowed consecutive transmissions of frames from the same access category without the need for backoff by using spacing between consecutive frames of SIFS, which is the shortest inter-frame spacing. The station thus maintains control of the channel for the entire TXOP by waiting a shorter time between transmissions than any other station. A good portion of a TXOP is also protected from collisions with hidden terminals. The acknowledgement to a frame indicates in the Duration field the length of the following frame in the TXOP or the remaining TXOP duration. This length is derived from the Duration field of the transmitted frame that is acknowledged. Because all but one frame in a TXOP is transmitted without contention, TXOPs help reduce the frequency of collisions. This increases channel use efficiency.

Another useful efficiency measure introduced in 802.11e is the expiration of frames based on the time queued, waiting for transmission. The time limit for the expiration of a frame, known as MSDULifetime, varies with access category, as overly delayed frames may not be useful in some applications but useful in others. For instance, applications with low latency tolerance, like voice, drop excessively delayed frames on the receiving end. Excessively delayed frames are dropped in this case without further transmission attempts, thus making room for other transmissions. Naturally, one must be careful when setting MSDULifetime limits for different access categories to separate traffic types with different tolerance for packet loss. For instance, VoIP signaling and VoIP media should be in separate access categories in such a case.

The impact of dropping excessively delayed frames has been studied in [5]. Figure 3.1 shows the effect of dropping voice frames if the time spent in the MAC layer exceeds the MSDULifetime limit. The average over-the-air delay experienced by a single station engaged in a voice call without dropping frames appears in Figure 3.1(a). Figure 3.1(b) shows the same resulting delay as a consequence of dropping any frames delayed by 20 milliseconds.

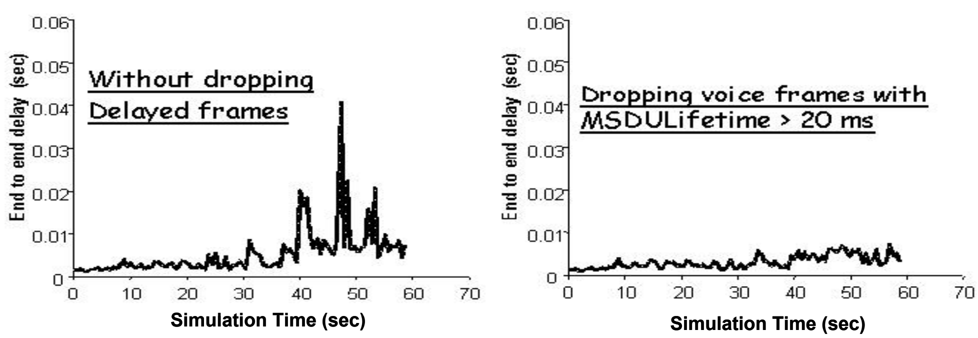

Figure 3.1: Effect of dropping frames delayed in excess of MSDULifetime.

It is important to note that, because the contention-based access mechanism of 802.11e is based on access prioritization, part of the advantage it brings to a WLAN over the legacy CSMA/CA protocol is lost when the traffic load consists primarily of one type of traffic – as for instance in the case of call centers, where much of the traffic comprises voice calls. The channel overhead penalty introduced by the longer MAC headers of 802.11e frames may counteract efficiency gains introduced in EDCA. The benefit of choosing EDCA over legacy DCF in such a case relates to the greater flexibility found in the former, as illustrated in Section 3.4.

3.2.2.1 TCMA MAC Protocol

According to the CSMA/CA protocol, as implemented by 802.11, a station engaged in backoff countdown must wait while the channel is idle for time interval equal to DIFS before decrementing its backoff immediately following a busy period, or before attempting transmission. According to the TCMA (Tiered Contention Multiple Access) protocol, variable lengths of this time interval, which is called Arbitration-Time Inter-Frame Space (AIFS), lead to varying degree of accessibility to the channel [5,6]. A shorter AIFS will give a station an advantage in contending for channel access. Differentiation between different access categories is achieved by assigning a shorter AIFS to a higher priority access category. An example is shown in Figure 3.2.

The effectiveness of priority differentiation of access categories is only partly due to allowing the station with the shortest waiting requirement to access the channel first, given two or more stations with expired backoff. This mechanism was used in 802.11 to give priority to stations engaged in PCF to access the channel before any other station. For instance, an AP would wait an idle time period of length PIFS, which is shorter than the length of DIFS required of a station. When a legacy AP has to engage in backoff, however, it uses the same backoff countdown rules as a station. It must wait for an idle interval of DIFS duration before decrementing its backoff timer.

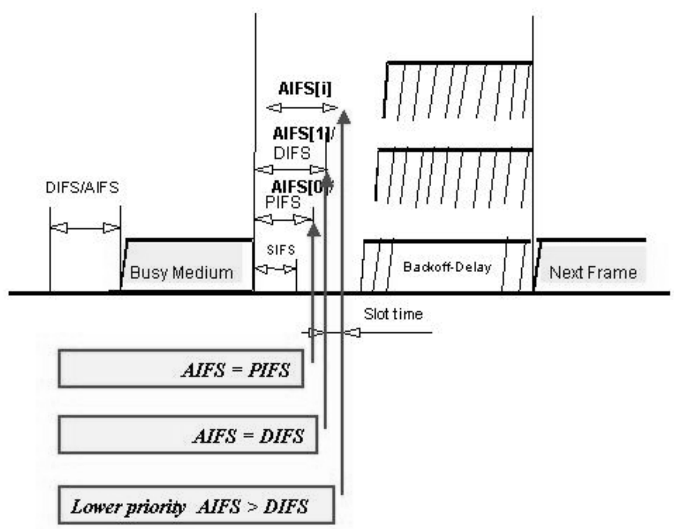

Figure 3.2: AIFS differentiated contention-based access.

The benefit from shortening the waiting time for transmission is small relative to the effect of the different AIFS values when decrementing the backoff timer. Since countdown of the backoff timer following a busy period may not occur unless the channel has been idle for a time period equal to AIFS, backoff countdown of lower priority frames slows down, and even freezes, in the presence of higher-priority frames with expired backoff. This is because a transmission will occur and the channel will be busy again before the lower-priority station, with the longer AIFS, has a chance to decrement its backoff timer. This would occur commonly in congestion.

Hence, in congestion conditions, the priority mix of stations with expired backoff timers favors higher priorities. In general, high priority stations will have lower backoff values than lower-priority stations when one looks at the residual backoff values of a mix of stations at any point in time. This desirable result is achieved without shortening the contention window from which the backoff value is drawn, which if pursued would increase the likelihood of collisions among the high-priority stations. Given any mix of initial random backoff values, the tendency of high-priority frames to reduce their backoff faster than lower-priority frames under TCMA leads to lower delay and jitter than without AIFS differentiation. Finally, the same tendency also reduces the likelihood of collisions between frames of different priorities, thus leading to a lower collision rate and higher throughput. These observations, which lead to the adoption of AIFS differentiation into the 802.11e standard, have been confirmed by subsequent performance evaluation studies [11 – 12].

3.2.3 802.11e Polled Channel Access

The IEEE 802.11e standard improved the PCF polled channel access mechanism of the earlier 802.11 standard to achieve better delay and jitter performance and greater channel use efficiency. The enhanced mechanism, called HCF controlled channel access (HCCA) in the 802.11e standard, resembles PCF, but with the following modifications. Polling is not limited to the contention-free period, but instead it can occur any time. The polling schedule is tailored to the time profile of the individual traffic streams, thus reducing both overhead, delay, and jitter. Overhead, delay, and jitter are also reduced through uplink TXOPs, which cause frames to be transmitted sooner than would have been possible otherwise.

In general, a *service period* is a time interval of continuous communication between the AP and a station, comprised of downlink transmissions and/or a poll and the station’s response to the poll. Polled-access service periods occur periodically at a negotiated service interval subject to limited time slippage. The AP transmits downlink frames to stations as single frames or as TXOPs. A downlink frame may be combined (or piggybacked) with a poll. With the poll, the AP grants a polled TXOP to the station. That means a response to a poll may consist of multiple uplink frames. An uplink frame can be combined with the acknowledgement to a downlink frame. The station can request extension of the TXOP by indicating the desired duration in a special QoS control subfield: TXOP Duration Requested. Uplink transmissions are protected from contention from other stations in the WLAN for the value of the Duration subfield in the downlink frame(s) sent to the station during the station’s service period.

By allowing multiple frames to be transmitted uplink without contention, in response to a single poll, a lot of the signaling frames that would otherwise be required are eliminated. TXOPs reduce contention when employed by either access method. TXOPs that are secured by the AP and granted to a station employing polled access give the station priority over any station using contention-based access, regardless of their respective priorities.

To match polling frequency to the traffic, a station that starts a new traffic stream exchanges signaling with the AP to establish the schedule by which the station will be polled. A station may have several traffic streams going on at once. An ADDTS frame is submitted for each traffic stream associated with the station, describing various aspects of transmission/delivery in the TSPEC element. These include the following: the nominal size of data frames (Nominal MSDU Size), the average bit rate at which data is generated (Mean Data Rate), the maximum delay allowed for queuing and transport of frames across the channel (Delay Bound), the maximum time allowed between consecutive service periods granted to the station (Maximum Service Interval) for the traffic stream, and the minimum physical bit rate to be assumed in establishing a schedule (Minimum PHY Rate). Each stream may have a different polling schedule. Alternatively, a station may request a single aggregate polling schedule for all admitted traffic streams. It does so by setting the Aggregation subfield in the TS Info Field of the TSPEC element equal to 1.

If the AP can accommodate the stream specified in the ADDTS request, it will indicate so in an ADDTS response that includes the Schedule element, specifying the schedule of the delivery of data and polls. If an ADDTS request is declined, the station may employ contention-based access for the traffic stream. A traffic stream is deleted when a station sends a DELTS frame to the AP. The negotiation between the station and the AP in establishing a polling schedule for each traffic stream, through the submission of ADDTS frames, provides a stand-alone admission control mechanism. As explained above, polled access has priority over contention-based access. It is not necessary, therefore, to restrict access of coexisting contention-based stations through admission control in order to enable polled stations to enjoy guaranteed delay/jitter performance.

The enhanced polled access mechanism of the 802.11e standard may operate during both the contention and the contention-free periods into which the channel time is typically partitioned. The AP can access the channel during the contention period by using PIFS, a shorter waiting requirement than that for stations, to initiate service periods for the stations with admitted traffic streams [13 – 15]. As a consequence, it is expected that, in practice, 802.11e APs will allocate most of the channel time to contention periods.

Compared to the legacy PCF mechanism, the 802.11e polled access mechanism results in a polling schedule that better matches the generation of frames in a periodic traffic stream. This results in superior delay/jitter performance and better channel use efficiency. The transmission of multiple uplink frames per poll also increases channel use efficiency.

Relative to contention-based access, scheduled polled access leads to better channel use efficiency because stations in the same WLAN (that is, stations served by the same AP) do not contend for the channel, thus eliminating the possibility of collision among them. The superior delay/jitter performance of polled access in 802.11e makes it the ideal choice for voice and streaming multimedia applications.

3.2.4 Illustrative Examples

Time-sensitive traffic occurs in diverse environments, with a different mix of traffic priorities. The prioritization capability of EDCA has been demonstrated in several performance studies [5, 9, 12]. Figure 3.3 illustrates the impact of AIFS differentiation on the average over-the-air delay experienced by nine high-priority voice streams using an 802.11b channel in the presence of lower priority data traffic, considered in [5]. Figure 3.3(a) shows the average delay experienced by the voice streams if the legacy DCF access mechanism was used and Figure 3.3(b) shows the delay experienced with EDCA.

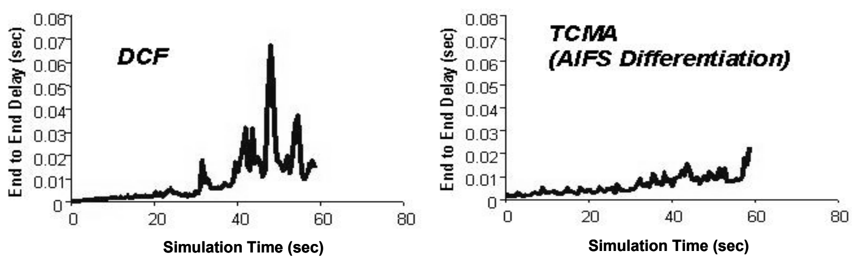

Figure 3.3: Average delay for top priority traffic category.

Prioritized access is useful if both low and high priority traffic are present in the same WLAN. The question thus arises whether EDCA would be of value in WLANs carrying mostly traffic of the same priority, such as call centers. The value of EDCA in such environments stems from its flexibility and the efficiency of channel utilization it introduces. For instance, EDCA can be of benefit because it allows the AP to use different access parameters than the stations.

Identical EDCA access parameters across all entities contending in a given priority class lead to consistent performance for all the traffic in that priority class only if these entities have comparable traffic loads. There is a pronounced load-induced inequity in the case of the AP. The AP has more traffic to transmit than any individual station since the uplink traffic is distributed among multiple stations and, in general, the downlink traffic in a WLAN is heavier than the uplink traffic. In the case of voice calls, the AP must transmit multiple voice streams, one for each station engaged in a voice call, while the stations transmit one voice stream each. By allowing the AP to contend for the channel with higher-priority EDCA parameters, downlink delays are shortened and become comparable to those of uplink voice streams.

Allowing the AP to access the channel with a shorter AIFS duration than the stations and no backoff requirement increases the voice capacity of a WLAN by as much as 38 per cent [16]. The voice capacity of a WLAN is the number of simultaneous voice calls that result in bounded over-the-air delays and no buffer overflow. Assuming an error-free channel, 46 voice calls with 20 milli-second frame interarrival time can be carried in an 802.11a WLAN when the AP uses the same access parameters as the stations. The WLAN voice capacity becomes 58 when the AP is allowed to transmit with AIFS equal to PIFS and a contention window of size zero.[1]

[1] Since the conference proceedings where the results in reference 16 of this chapter are not readily available, a synopsis is included as an endnote.

Using the HCCA polled access mechanism of 802.11e can increase the voice call capacity of a WLAN further, as it provides collision-free transmission. For comparable conditions, the voice capacity of an 802.11a WLAN is 65 voice calls [17].

The use of HCCA introduces a 12 per cent increase relative to the capacity achieved with optimized AP access parameters. Such a gain may seem insufficient to justify the complexity of implementing the scheduling algorithm required for HCCA. Considering EDCA as the alternative, some algorithmic complexity is also needed in order to achieve high capacity consistently. It relates to the choice of the access parameter values for different traffic conditions. The 802.11e standard does not specify how these parameters must be set; a task left to the user. The appropriate choice of a certain EDCA parameter value – namely, the contention window size – depends on the traffic conditions. The wrong choice could result in capacity loss, because of aggressive behavior and a high collision rate. This point is illustrated in [17], where choosing the standard default contention window value leads to a capacity of only 35 voice calls for an 802.11a WLAN employing EDCA. In the absence of special optimization algorithms for adaptation of the contention window size to traffic intensity, this EDCA parameter should be assigned a large fixed value.

Naturally, assigning large fixed values to the EDCA contention window size removes its effect on prioritization, leaving the AIFS size as the main priority differentiator between traffic classes. While this works for exclusively 802.11e WLANs, mixed systems are problematic. The range of ten time slots provided in the 802.11e standard for the AIFS duration is sufficiently wide to enable adequate priority differentiation [11, 12]. The entire range is not available, however, when 802.11e-compliant stations must co-exist with legacy stations. Legacy stations, which employ an AIFS interval of length DIFS+1, must be treated as having low priority traffic. The effective AIFS range is thus reduced to a single time slot, which may not be sufficient for differentiation among multiple classes. Additional differentiation would thus be of value. Hence the need to differentiate based on contention window size in this special case, in spite of the caveats.

3.3 Admission Control

Admission control provides bandwidth management to ensure that QoS-sensitive applications, such as voice and video, will be afforded a satisfactory quality of service. Overloading the WLAN with an excessive number of users entitled to high-priority access would make it hard to provide consistent QoS. Therefore, requests are submitted by stations for the admission of specific traffic streams to the AP, which keeps track of the traffic on the channel and accepts or declines the request. The information contained in this exchange will depend on the channel access method involved; it will be different for contention-based access and for polled access.

Admission control is an intrinsic part of polled access, and thus comes automatically with the decision to use this access method. Admission control is an option that is available for stations using contention-based access. It is important to note that admission control becomes imperative for contention-based stations with QoS traffic in a WLAN that supports polled access, unless polled access is limited to just top priority traffic. Stations using contention-based access will access the channel with lower priority than any station that uses polled access, regardless of their respective traffic priorities. Because the AP can transmit before any station, it can give a polled station an opportunity to transmit before any contention-based station.

3.3.1 Admission Control for Contention-Based Channel Access

Admission control for contention-based access is an optional feature for a station and an AP. It involves the decision at the AP to allow stations that employ contention-based access in the WLAN to transmit traffic using the parameters of an access category. This enables the AP to track and manage bandwidth use. It is not necessary to impose admission control on all access categories. The 802.11e contention-based access mechanism shields the admitted traffic from contention by lower-priority transmissions. It is important, however, to require admission control in all access categories of higher priority than the access category of the traffic of interest. The contention to be experienced by traffic in a given access category cannot be bounded if traffic in access categories of higher priority is unrestricted.

The basic procedure of admission control for distributed access is the following. In its beacons, the AP advertises to the WLAN the access categories that are protected by admission control. A station that has traffic to transmit or receive in a protected access category must request permission from the AP before it is allowed to do so. The signaling is similar to that used for the admission of a traffic stream for polled access [18]. A station’s request, submitted in an ADDTS (add Traffic Specification) frame, describes the ‘traffic stream’ to be admitted. The description includes the data frame size, the mean data rate, and the minimum physical transmission rate for each of the directions on which the channel would be accessed with the parameters of the access category in question. If an ADDTS frame indicates a bi-directional traffic stream, traffic is specified for one of the two directions; the other is assumed to be the same.

The response to the ADDTS request, if affirmative, furnishes in the Medium Time field the ‘channel time’ the station is allotted for uplink transmissions using the parameters of the access category specified in the request. The allotted channel time is expressed as the number of time units the channel may be used by the station for its transmissions over a fixed known time interval. If the AP declines an ADDTS request, the station may still transmit, but with parameters of a lower-priority access category that requires no admission control. There should be at least one access category without the admission control requirement. Stations that do not support admission control may transmit only with parameters of access categories of equal or lower priority, and for which admission control is not mandatory.

Once a station receives its allotted channel time for a particular access category, it keeps track of the portion that has been used up for its transmissions, and for any retransmissions. The station may request additional channel time for an admitted traffic stream if its allocation is being used up too fast, or if a new data flow is added to the same traffic stream. A single admitted traffic stream could be specified per access category, which would be the aggregate of several data flows. The station updates the combined requirements of all data flows in the access category in question and sends a new ADDTS request for an updated allocation. To give up all of its allotted channel time for a particular access category, a station submits a DELTS (delete Traffic Specification) frame. The channel time allotted to a station for an access category is released if no transmissions in that access category to/from the station have occurred for a specified time period, the Inactivity Interval, which is indicated on the ADDTS frame.

3.3.2 Admission Control for Polled Channel Access

Admission control is exercised automatically when using polled channel access. The AP will reject an ADDTS request if it cannot meet the requirements for a service period schedule requested by a station for a traffic stream. If the requested requirements can be met, the AP responds with a service period schedule. Unlike in the case of contention-based access, a station using polled access may have several admitted traffic streams of the same priority.

During the negotiation, a minimum set of parameters must be specified in the ADDTS request so that the AP can schedule time on a service period for the traffic stream that is to be admitted. These parameters include mean data rate, frame size, minimum transmission rate, and either the maximum service interval or a delay bound. If a traffic stream is admitted, the ADDTS response will include non-zero values for mean data rate, frame size, minimum transmission rate, and the maximum service interval. The ADDTS response will include a Schedule element, which provides the schedule of the delivery of data and polls. The minimum transmission rate will be used in determining the length of TXOPs and service periods.

The priority of a traffic stream may be considered in admission control. An admission control request from traffic stream with a higher priority may cause an admitted stream to be dropped. The AP sends a DELTS frame to notify a station that a traffic stream is dropped. Admission of a traffic stream may therefore not be guaranteed.

3.4 Power Management

Several of the QoS-sensitive applications will involve multimedia traffic over battery-powered handheld devices, such as a PDA or a wireless VoIP phone. In crafting a standard of good QoS performance, it was thus considered important to prolong the battery life of such devices. The 802.11e standard amendment offers several new mechanisms to help battery-powered devices conserve power by enabling them to power down their receivers and transmitters intermittently without losing connectivity or data. The new power management mechanisms apply to WLANs served by an AP – such WLANs are known as 'infrastructure' WLANs, and for this reason, the discussion in this section will focus on power saving methods for infrastructure WLANs.

A station informs the AP of its operating power-management mode, 'power saving' versus 'active', when it associates with the WLAN. The mode can be changed during the association period by changing the Power Management bit, a bit in the frame control field of the frames transmitted by the station. The AP will not send transmissions to a station that has declared itself to be in 'power save' mode, unless it knows that the station has its receiver fully powered, i.e., it is in the 'awake' state, and ready to receive. Otherwise, the AP will assume that the station's receiver is powered down, i.e., it is in the 'doze' state, and for this reason, any incoming frames addressed to a power-saving station will be buffered for later transmission.

A simple, but not efficient, way for a power-saving station to retrieve multiple buffered frames at once is to switch its power management mode to 'active'. A data frame, or a Null frame sent by station to the AP with the Power Management bit set to 0 will enable the AP to transmit the buffered data. The station may subsequently return to power-save mode using another frame with the Power Management bit set to 1. The inefficiency in this approach stems from the fact that it requires extra frames to be transmitted for signaling purposes. 802.11e introduces delivery methods with reduced signaling.

The AP may deliver buffered frames to their destination power-saving stations either on a previously negotiated schedule or in response to transmissions from the respective stations that initiate such delivery. In order to initiate delivery in the latter case, a station should know that there are frames for it buffered at the AP. Notification of the presence of buffered frames at the AP typically comes through a special station-specific field contained in the beacon frames broadcast by the AP, or in reserved fields of downlink frames directed to the individual stations. In some situations, as we will see, notification is not provided by the power-saving mechanism, and thus it must be furnished either by the application running on the station, or by transitioning to a different power saving mechanism that provides such notification.

The station chooses the delivery and notification mechanisms and communicates it to the AP either upon association or re-association of the station with the WLAN or through explicit signaling using an ADDTS frame. The various mechanisms available in an infrastructure WLAN will be described in the following section. They include (1) the 'legacy' power save mechanism, which was available pre-802.11e. APSD (automatic power save delivery) was introduced by the 802.11e standard to reduce the signaling that would otherwise be needed for delivery of buffered frames to power-saving devices by an AP. APSD provides two ways to start delivery: (2) 'scheduled APSD' (S-APSD) and 'unscheduled APSD' (U-APSD). Unscheduled APSD can take (3) a 'full' U-APSD or (4) 'hybrid' U-APSD form. With full U-APSD, all types of frames use U-APSD independently of their priority. Hybrid U-APSD employs a combination of U-APSD and the legacy power save mechanism.

3.4.1 Legacy Power-Save Mechanism

The legacy power-save mechanism applies to both infrastructure WLANs and WLANs without an AP. We describe here how it works with the former since the new power save mechanisms deal only in WLANS served by an AP. For information on the latter, the reader is referred to the 802.11 standard.

Frames buffered at the AP for a power-saving station employing contention-based access are delivered when the station sends a special control frame, the power save poll (PS-Poll). The AP sends a single buffered frame to a station after receiving a PS-Poll, either immediately or soon thereafter. More PS-Polls are required in order to retrieve additional buffered frames. The presence of further frames remaining at the AP is indicated by the More Data bit of the control frames of the transmitted frame, which is set to 1.

A station using legacy power save can rely on the traffic indication map (TIM) to learn if the AP holds buffered data for it. The TIM is a bit map containing the buffer status per destination station. It is sent regularly on beacon frames broadcast by the AP at known times. If the station is in the ‘doze’ state, it will wake up at the beacon times to receive and interpret the TIM. Alternatively, a station can ascertain the presence of additional frames buffered for it at the AP while receiving a buffered frame. The More Data bit in the control field of that frame would have been set to 1 if additional frames remained buffered for the station.

A power-saving station that supports legacy polled access need not send PS-polls in order to receive its buffered frames. The station receives its buffered frames at the start of the contention-free period, when it awakens to listen to the TIM and learn of its buffer status at the AP. Such a station would probably not request to be on the polling list because that would require staying awake for the entire contention-free period. Uplink frames are sent by contention in that case.

3.4.2 Automatic Power Save Delivery

APSD is a mechanism for the delivery of unicast frames from the AP to a power-saving station. This mechanism was introduced by 802.11e in order to reduce the signaling traffic caused by PS-Polls and their acknowledgements. A station may use both APSD and legacy PS-Polls at the same time to retrieve buffered frames from the AP. Certain restrictions apply, however, which are discussed below. To use APSD, stations must have the Power Management subfield in the control field of all transmitted frames set to 1.

The AP may deliver buffered frames to their destination power-saving stations either on a previously negotiated schedule or in response to receiving transmissions from the respective stations that trigger such delivery. The two APSD approaches are thus known as ‘scheduled’ and ‘unscheduled’. A station may use both approaches at the same time, provided that only one is used for a given access category.

3.4.2.1 Scheduled APSD

This mechanism is well suited for periodic traffic streams, such as voice and audio/video, and is especially good for unidirectional downlink periodic streams. With scheduled APSD, downlink transmissions to power-saving devices will occur at a schedule that is known in advance, obviating the need for special signaling between the station and the AP.

The AP and the station negotiate in advance a time schedule by which the station will power its receiver fully to receive any frames that are buffered for it at the AP. A station that wishes to use S-APSD must send an ADDTS request with the APSD subfield in TS Info field of the TSPEC element set to 1. The TSPEC element contains the time of the first downlink transmission (Start Service Time) as well as the time interval at which downlink transmission will be repeated (Service Interval), as in the case of polled access. The Start Service Time is expressed in terms of the time shared in the WLAN, known as the TSF timer [19]. While the Start Service Time field is used optionally with polled access, this field must be specified when using Scheduled APSD, as knowledge of the time of downlink frame delivery affords a station the longest stay in the doze state.

The AP is given the last say in setting the start time of the periodic transmissions to the station so that its transmissions to different power-saving stations are staggered in a way that minimizes the time the power-saving stations are awake. If the request is accepted, the AP will return an ADDTS response containing a Schedule element, which, among other, includes the Start Service Time selected by the AP. The station will wake up to receive its buffered frames at the times indicated by the returned schedule.

Either channel access method, polled or contention-based access can be used with Scheduled APSD. Scheduled APSD fits naturally with polled access. To indicate polled access, the Access Policy subfield of the ADDTS TS Info field would be set to (0, 1), and the Start Service Time field in the TSPEC element must have a nonzero value. When the station plans to use contention-based access with Scheduled APSD, the Schedule subfield of the ADDTS TS Info field must be set to 1, and the Access Policy subfield must be set to (1,0).

For stations using Scheduled APSD in conjunction with contention-based access, the uplink transmissions do not require polling. A power-saving device that uses contention-based access can transmit to the AP at any time.

3.4.2.2 Full Unscheduled APSD

Unscheduled APSD was introduced for stations accessing the channel by contention, in order to enhance the efficiency of legacy power save. A power-saving station may use not just a PS-Poll, but also any data or Null frame – referred to as a 'trigger' frame – in order to notify the AP that its receiver is fully powered and ready to receive transmissions [20, 21]. Using a data frame that is pending transmission at the station, instead of a PS-Poll, to initiate downlink transmission clearly reduces the traffic generated by the station and increases battery life and channel use efficiency.

Additional gains are achieved from relaxing the number of frames the AP is allowed to transmit to a power saving station when it receives notice to do so. While receiving a PS-Poll from a station allows the AP to transmit a single downlink frame -- of the highest priority access category buffered -- receiving a trigger frame will start an APSD service period for that station. During a service period, the AP may send multiple frames, subject to a limit specified by the station. Eliminating the extra signaling that would otherwise be necessary under legacy power save also increases the efficiency of channel use and conserves battery life.

Naturally, since the station does not know in advance the number of frames sent by the AP in a service period, it must be notified when the last frame has been transmitted for a given service period so it may transition to the doze state. The control subfield EOSP in the last delivered frame marks the end of a service period.

The AP need not deliver all frames buffered for a station in a single service period. As in the case of legacy power save, the More Data control subfield in a last frame transmitted in a service period indicates whether there are frames remaining buffered at the AP. Knowing its AP buffer status enables the station to send another trigger frame or PS-Poll to retrieve more of its buffered frames.

As with the legacy power-save mechanism, a station can learn about its buffer status by listening to the beacons for its TIM [21]. This is needed only while not receiving frames from the AP, as the More Data control subfield in downlink frames to the station conveys the same information.

To use full U-APSD, a station sets the first four bits of the QoS Info subfield of the QoS Capability element in the (re-) association request all to 1. The Max SP Length subfield of the QoS Info field is used to place a limit on the maximum number of frames to be delivered during a service period. For unrestricted delivery, this subfield should be set to 0.

3.4.2.3 Hybrid Unscheduled APSD

The hybrid U-APSD mechanism allows a station to choose between legacy power save delivery and APSD based on access category [22]. Trigger frames are used to initiate the delivery of buffered frames associated with access categories that have been designated as ‘delivery enabled’. Buffered frames of access categories not so designated, can be retrieved with PS-Polls only. The station also designates in advance the access categories of the frames that may serve as trigger frames. An AP receiving frames in categories other than those designated by a station as trigger-enabled will not transmit buffered frames to the station.

The end of a station’s service period and the presence of further frames remaining buffered at the AP are indicated by the control subfields EOSP and More Data of frames received by the station. However, unlike in full U-APSD where the More Data bit indicates the presence of buffered frames remaining at the AP, the same bit in hybrid U-APSD would indicate only whether frames of similar characterization (e.g., delivery enabled versus non-delivery enabled) as the received downlink frame remain buffered.

One can visualize the hybrid U-APSD mechanism as partitioning the incoming traffic of the different access categories into two sets, each directed to a different power-save buffer for a station, the legacy and APSD buffer. The notification and retrieval mechanisms work independently of one another, with PS-Polls used to retrieve frames from the legacy buffer, and a trigger frame causing frames to be delivered from the APSD buffer. The More Data bit on a downlink frame shows the status of the buffer in which the frame was held.

While stations receiving buffered frames know whether additional frames remain buffered from the More Data control subfield, stations not receiving any frames may have no way to knowing that frames are waiting at AP. The TIM in the beacons is used differently in hybrid U-APSD mechanism than the other power save mechanisms. It shows whether the AP has buffered for the station frames of non-delivery-enabled access categories only. Since the TIM does not account for frames of delivery-enabled access categories, a more pro-active way is needed for a station to figure out whether it has buffered frames of such access categories and must therefore initiate frame retrieval.

A station that has no uplink traffic may send a Null frame of a trigger-enabled access category uplink in order to both check buffer status and, if frames are buffered, initiate their retrieval. The AP responds with a Null frame if there are no frames buffered in the delivery-enabled access categories. The spacing of the uplink Null frames cannot be too long if the traffic in the delivery-enabled access categories is delay sensitive. The frequency of such Null frames cannot be high either, as that would increase channel load and battery drain unnecessarily.

Sending Null frames in order to retrieve buffered frames is inefficient for a power-saving station if it does not generate regular uplink traffic in a trigger-enabled access category and does not expect regular downlink traffic in a delivery-enabled access category. A more efficient way to retrieve buffered traffic in such conditions is to alter the characterization of the delivery-enabled access categories so that the status of the corresponding buffers will be included in the station's TIM. There are two options. One is to disable automatic delivery in all access categories, and then retrieve frames one at a time with PS-Polls. The second option is to enable automatic delivery for all access categories, and retrieve frames using full U-APSD. Changing the characterization of an access category requires further signaling – that is, re-association or the submission of TSPEC requests, one for each affected access category.

The designation of access categories as delivery- or trigger-enabled occurs through (re) association frames, by setting the corresponding subfields in the QoS Info subfield of the QoS Capability element. These designations may be also set or altered for an access category by submitting ADDTS frames for that access category, one for downlink and another for the uplink direction, indicating the new delivery and triggering capabilities, respectively.

To unify signaling for full and hybrid U-APSD, the convention was adopted to have a station employing full U-APSD designate all its access categories as delivery enabled.

3.4.3 Illustrative Examples

Examples of the use of APSD in various applications are given in this section. Aside from technical restrictions, some application can benefit more from a specific choice of a power-save mechanism.

Examples of an application for which Scheduled APSD provides an ideal power save mechanism are one-way periodic streams, like Internet Radio. In such applications, with acknowledgements suppressed, the traffic load consists primarily of periodic unidirectional streams from the AP with occasional uplink frames. Scheduled APSD enables the transmission of the periodic stream without the need for redundant uplink transmissions.

Scheduled APSD used in conjunction with contention-based access enables any frames generated by the station to be transmitted immediately without waiting for the station to be polled. In a congested WLAN, experiencing a lot of uplink delay jitter, Scheduled APSD prevents downlink traffic from assuming this delay jitter from uplink frames, as would be the case with U-APSD. Hence, Scheduled APSD would result in lower delay.

Finally, Scheduled APSD used in conjunction with contention-based access is ideally suited for wireless ad-hoc and mesh networks. In such networks, the power-saving device wakes up to receive frames that have been stored by a neighboring device according to a pre-negotiated schedule.

Full U-APSD is a simple, efficient, power-save mechanism appropriate for any mix of traffic, uplink and downlink, periodic and non-periodic. With bi-directional periodic streams, as for example a wireless phone application involved in a call, traffic flows back and forth between the station and the AP at regular intervals. Frames on the uplink stream cause the delivery of the downlink buffered frames. The More Data control subfield of the downlink frames notify the station as to whether more frames remain buffered at the AP. After receiving a portion of its buffered traffic, the station thus knows it must pursue further retrieval of the remaining frames. It is sufficient, therefore, for a voice-enabled device to listen for its TIM only while on standby in order to receive notice of the presence of any buffered frames, from any application, including signaling for incoming calls. Not having to listen for the TIM preserves the battery of the device.

Hybrid U-APSD offers a power saving device the ability to control the time used for the delivery of various types of traffic. This is useful when the device must tend to other time-critical activities that require postponement of the delivery of some types of traffic. Such activities would include a wireless phone engaged in channel scanning other channels for roaming possibilities. By restricting APSD delivery and triggering only to the top-priority access category during a call, a station engaged in call will receive from the AP only its buffered voice frames when it transmits voice frames uplink. However, the duration of power-save delivery can also be controlled directly by setting the Max SP Length field in the station's association request to the desired duration for a service period.

A VoIP-enabled station using hybrid U-APSD must take measures to receive notification of the arrival at the AP of any out-of-band signaling, or other traffic, regardless of whether this traffic is associated with a delivery-enabled access category or not. The station awakens to listen to the TIM in order to receive notice of buffered traffic not associated with a delivery-enabled access category. As for buffered voice frames at the AP, notification is received with the delivery of downlink voice frames in the frame's control field. When the call ends, however, there are no voice frames to convey the station's buffer status for the delivery-enabled categories. So, when the station goes on standby, it changes the delivery characterization of its access categories to non-delivery enabled in order to receive notice of their buffering through the TIM.

3.5 QoS in Wireless Mesh Networks

The concept of a wireless mesh takes on several forms, the most common being a collection of nodes that form an ad hoc network and are capable of serving as WLAN APs. Using the wireless channel, these nodes, which are called 'mesh points', can forward traffic received from 802.11 stations to other mesh points with ultimate destinations that include WLAN stations attached to other mesh points or points somewhere on the wired network. In addition to the 802.11 standard, protocols for forwarding, routing and channel access must be specified for the mesh points.[2] This requires an ad hoc networking standard with multi-hop capability.

A wireless mesh network can be used to enable WLAN service when wiring for APs is not readily available in an enterprise, or for a temporary network that can be easily set up and torn down. Mesh networks have found applications in public safety, disaster control, surveillance, and connectivity for municipal services. Municipalities and service providers are interested in wireless mesh for providing public access, an alternative to expensive home broadband in dense urban areas, or to offer inexpensive WiFi service to rural communities. In some applications, mesh points may be simply stations communicating wirelessly on a multi-hop ad hoc or infrastructure network.

The wireless mesh presents multiple challenges, including challenges in routing and security, especially when mobility is contemplated. While it is clear that tools exist to handle these issues, it is not clear what choices will ultimately be made for the IEEE 802.11 mesh standard. Channel selection, channel access, and meeting QoS requirements also present challenges in a wireless mesh network.

---

[2] The IEEE 802.11 Task Group s is developing a standard for 802.11 mesh networks.

The wireless medium providing forwarding and backhaul service for the mesh points may use either 802.11 connections, in the unlicensed spectrum used by WLANs, or some other wireless technology operating in different RF bands. Using 802.11 technology, while keeping costs low and connection simple because of the unlicensed RF spectrum, presents challenges arising from the competition between WLAN and mesh traffic. Dedicating different radios to the two types of traffic, each operating in one of the two unlicensed RF bands or on non-overlapping channels of the same RF band, eliminates the competition between WLAN and mesh traffic. By allowing these two types of traffic to be served by the same radio(s), however, and by properly managing the competition between them, one can also reduce hardware costs. RF management should be done in a distributed manner, as the requirement for controllers for this purpose might prevent hardware of different vendors from being interoperable.

Part of distributed RF management involves channel selection and channel access. Any group of mesh points that can hear one another must be able to operate on multiple channels in order to increase the mesh traffic carrying capacity. The value of single channel meshes is mostly in the home or small office where the total traffic does not exceed the traffic that can be carried by a WLAN. Single channel mesh points are useful also when there is need to set up communication quickly over a large area without wiring, as for instance in disaster control. The purpose of a single-channel mesh is primarily to extend coverage range of a WLAN through multiple-hop transmissions. As the number of hops increases, however, there would be a decrease in the mesh goodput (that is, the amount of successfully transmitted traffic that originates or terminates in the mesh), because more channel time is taken up to transmit frames end-to-end.

Another challenge is managing latency in order to meet QoS requirements. The latency experienced by frames traversing a wireless mesh increases fast when traffic bottlenecks arise as a result of traffic concentration in parts of the mesh that lack the necessary throughput capacity. For instance, this occurs in access mesh networks, where traffic concentrates at nodes near the point of attachment of the mesh to the wired network. If these nodes are not equipped with multiple radios in order to handle the higher throughput, they will become bottlenecks, contributing to frame delays and frame loss. Single radio mesh points are inadequate for a wireless mesh used to connect multiple APs. Use of two radios per node for the sole purpose of separating WLAN traffic from mesh traffic, though helpful in reducing the competition between them, is inadequate when the mesh comprises multiple APs.

As traffic is forwarded from node to node on a multiple-hop path of a wireless mesh, the latency experienced at each node adds. The 802.11e standard is useful in the wireless mesh where prioritization in a single hop can put time-sensitive traffic on the air before other traffic. However, 802.11e prioritization alone will not address the challenges arising in a multi-hop network due to accumulating latency. Methods for managing this latency are needed in order to meet the QoS requirements of real-time streaming applications such as VoIP. An example of such a mechanism is given below.

‘Express’ forwarding is a mechanism that illustrates how the 802.11e could be augmented in order to limit the latency accumulated over a multiple-hop path on a wireless mesh backbone. The transmitting node adds a set amount to the duration field of a QoS-sensitive transmission, while specifying that the transmission is using the express-forwarding mechanism. The immediate destination node knows that it can transmit that set amount sooner than the end of the requested duration, while neighboring nodes respect the full request. Thus, subsequent hops are spared contention delays.

The objective of express forwarding is to reduce access delay experienced by forwarded traffic after the first hop. The first hop experiences an access delay similar to any single-hop transmission using the same EDCA access parameters. A mesh point that is forwarding voice traffic beyond a single hop would not experience additional access delay, as it is allowed immediate contention-free access to the channel. Having the Duration field of the frame transmitted on the first hop set to a longer value than is necessary to complete transmission silences neighbor nodes. The node to which the frame is forwarded is permitted to transmit sooner. Without contention from neighbor nodes, the receiving node can access the channel immediately and forward the frame on to the next hop without further access delays. The Duration field value is shortened only for frames sent to intermediate nodes of a multi-hop path, and not on the final hop.

3.6 Summary

The IEEE 802.11e standard offers QoS functionality at the MAC Layer. Several new features are introduced for channel access, admission control, and power save. With the new mechanisms, a WLAN will be able to differentiate between traffic of different priorities and provide faster channel access for higher-priority traffic. At the same time, the new standard also pursues more efficient ways of utilizing the channel and better power management techniques for battery-based stations.

As in the case of the 802.11 standard, the 802.11e amendment offers two forms of channel access, polled access and contention-based access. Admission control is an intrinsic part of the polled access mechanism (HCCA) but it is an available option for the contention-based access mechanism (EDCA). Power save is an independent capability from channel access. However, when a station pursuing scheduled APSD uses polled access, it behaves very much as if it uses HCCA. Scheduled APSD can also be used with EDCA.

3.7 Endnote

The results in [16] were derived from a performance study where two access categories were used, Voice (VO) and Best Effort (BE), for voice and data respectively. Two scenarios were considered. In the first scenario, referred to as ‘EDCA’, the AP and the stations access the channel with the same access parameters for both traffic priorities. In the second scenario, referred to as ‘PIFS Access’, the AP uses different access parameters to transmit voice. With a shorter AIFS and no backoff, the AP can access the channel faster. Table 3.1 summarizes the differences in access parameters for the AP under the two scenarios. All other parameters were set at their default values indicated in the 802.11e standard [1].

Table 3.1: AP Access Parameter Values.

Parameters*	**‘EDCA’**	**‘PIFS’**
AIFSN[VO]	1	0
CWMin[VO]	15	0
AIFSN[BE]	2	2
CWMin[BE]	31	31

*AIFSN is the priority-dependent number of time slots that determines the AIFS length.

Tables 3.2 and 3.3 show the effect of ‘PIFS’ access on call capacity of a WLAN, how the capacity is impacted by traffic contending at a lower priority for the same channel, and the influence of voice packet aggregation prior to delivery to the MAC layer. For the latter, two RTP frame payload sizes are considered: one with 10 milliseconds and another with 20 milliseconds of audio data. These results are presented for the 802.11b WLANs (for a transmission rate of 11 Mbps) and 802.11a WLANs (for a transmission rate of 54 Mbps), respectively.

The presence of BE data traffic in the WLAN reduces call capacity. For the voice stream generating frames every 10 milliseconds, a capacity of 10 calls is possible in the 802.11b WLAN without any other traffic, as seen in Table 3.2. The capacity reduces to 6 calls when a station transmits data at 12.3 Mbps. For the 802.11a WLAN, the call capacity drops from 24 to 20 calls when introducing a data load of 12.3 Mbps, as seen in Table 3.3. Clearly, although EDCA expedites the transmission of higher-priority frames, it does not totally eliminate the competition for the channel from lower-priority frames.

Table 3.2: Call capacity of 802.11b WLAN.

Data traffic	10 ms audio		20 ms audio	
(Mbps)	EDCF	PIFS	EDCF	PIFS
0	10	11	18	19
0.5	9	10	17	17
2	7	10	12	17
12.3	6	10	11	17

Table 3.3: Call capacity of 802.11a WLAN.

Data traffic	10 ms audio		20 ms audio	
(Mbps)	EDCF	PIFS	EDCF	PIFS
0	24	33	46	58
0.5	23	33	44	58
2	23	33	43	58
12.3	20	33	41	58

‘PIFS’ access at the AP causes call capacity to increase. The benefit is greater in the presence of heavier data load, when the increased competition for channel by data transmissions leaves the AP at a greater disadvantage in accessing the channel than the voice stations. Specifically, the call capacity of the 802.11b WLAN increases from 6 calls to 10 calls (with a 10 millisecond voice interframe spacing) with ‘PIFS’ access when the data load is 12.3 Mbps, as seen in Table 3.2. The capacity goes from 18 to 19 calls with ‘PIFS’ access when no data traffic is present in the 802.11b WLAN. For the 802.11a WLAN, capacity goes from 22 to 33 calls without data traffic, and from 20 to 33 calls with a 12.3 Mbps data load (and the same audio payload), as seen in Table 3.3. 'PIFS' access at the AP brings robustness to competition from lower-priority traffic.

Finally, the effect of frame aggregation of voice traffic was considered. Increasing the payload size of RTP packets of a VoIP stream (thus reducing their arrival rate) causes the call capacity to increase, as both per call overhead and contention are reduced. In a voice-only 802.11b WLAN, the capacity goes from 10 to 18 calls when the RTP packet interarrival time increases from 10 to 20 milliseconds. A similar gain is experienced with 'PIFS' access, going from 11 to 19 in a voice-only 802.11b WLAN. The highest call capacity for an 802.11a WLAN is 58 calls under 'PIFS' access with a 20 millisecond RTP payload. Larger RTP payloads would not be advisable, as they add longer delay and jitter, affecting voice quality adversely.

3.8 References

[1] Amendment to Wireless LAN Medium Access Control (MAC) and Physical (PHY) Layer Specification: Medium Access Control Quality of Service Enhancements, IEEE Std. 802.11e, November 2005.

[2] IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical (PHY) Layer Specifications, ANSI/IEEE Std 802.11, 1999 Edition.

[3] L. Kleinrock and F. Tobagi, "Packet Switching in Radio Channels: Part I – Carrier Sense Multiple Access Models and their Throughput Delay Characteristics", *IEEE Transactions on Communications*, Vol. 23, No. 12, pp. 1400 – 1416, 1975.

[4] IEEE Standard 802.1D, "Media Access Control (MAC) Bridges," 2004.

[5] M. Benveniste, "Tiered Contention Multiple Access (TCMA), a QoS-Based Distributed MAC Protocol", *Proceedings of PIMRC*, Lisboa, Portugal, September 2002.

[6] M. Benveniste, "An Enhanced-DCF Proposal Based on 'Tiered Contention' Multiple Access (TCMA)", Doc IEEE 802.11-00/457, October 2000.

[7] M. Hoeben and M. Wentink, "Enhance D-QoS through Virtual DCF", Doc IEEE 802.11-00/361, October 2000.

[8] W. Diepstraten, "Distributed QoS resolution", Doc IEEE 802.11-00/361, October 2000.

[9] S. Mangold, et al, "Analysis of IEEE 802.11e for QoS Support in Wireless LANs", *IEEE Wireless Communications*, pp. 40 – 50, December 2003.

[10] S. Choi, et al, "Multiple Frame Exchanges during EDCF TXOP", Doc IEEE 802.11-01/566, January 2002.

[11] J. W. Robinson, and T. S. Randhawa, "Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function", *IEEE Journal on Selected Areas in Communications*, Vol. 22, No. 5, pp. 917 – 928, June 2004.

[12] Z. Tao and S. Panwar, "Throughput and delay analysis for the IEEE 802.11e Enhanced Distributed Channel Access", *IEEE Transactions on Communications*, Vol. 54, No. 4, pp. 596 – 603, April 2006.

[13] M. Hoeben and M. Wentink, "Suggested 802.11 PCF Enhancements and Contention Free Bursts", Doc IEEE 802.11-00/113, May 2000.

[14] R. Meier, "An Integrated 802.11 QoS Model with Point-controlled Contention Arbitration", Doc IEEE 802.11-00/448, November 2000.

[15] M. Fischer, "A Hybrid Coordination Function for QoS", Doc IEEE 802.11-00/452, January 2001.

[16] M. Benveniste and V. Dham, "VoIP Call Capacity of a 802.11 Wireless LAN Using EDCF Access", *Wi-Fi Voice Conference 2004*, Paris - France 25-28 May, 2004, http://pubs.research.avayalabs.com/pdfs/WiFiVoice2004.pdf.

[17] S. Shankar, et al, "Optimal packing of VoIP calls in an IEEE 802.11 a/e WLAN in the presence of QoS constraints and channel errors", *IEEE GLOBECOM*, Vol. 5, pp 2974 – 2980, November 2004.

[18] R. Meier, et al, "Uniform 802.11e Admissions Control Signaling for HCF and EDCF", Doc IEEE 802.11-02/678, November 2002.

[19] M. Benveniste, et al, "Proposed Normative Text for Simplified APSD", Doc IEEE 802.11-03/107, January 2003.

[20] Y. Chen, et al, "Power management for VoIP over IEEE 802.11 WLAN", *IEEE Wireless Communications and Networking Conference*, Vol 3, pp. 1648 – 1653, March 2004.

[21] M. Benveniste, et al, "Some Power-save changes in 802.11e Draft", Doc IEEE 802.11-04/584r2, May 2004.

[22] S. Wang, et al, "Simplifications and Enhancements to Unscheduled APSD (U-APSD)", Doc IEEE 802.11-04/694, July 2004.