IEEE 802.11e: QoS Provisioning at the MAC Layer

Yang Xiao, The University of Memphis

Abstract

This article introduces the emerging IEEE 802.11e standard to support quality of service at the medium access control layer. Both the contention-based and content-free centrally controlled channel access mechanisms are introduced by describing not only the MAC protocol operations and parameters, but also the call admission techniques and scheduling algorithm that have been designed for IEEE 802.11e. Finally, we provide simulation results aimed to highlight the capability of the EDCF to differentiate the traffic classes.

Introduction

Net access through hotspots at airports, hotels, and coffee shops, via the high-speed wireless Internet access service known as WiFi, is rapidly becoming common. Recently, AT&T, IBM, and Intel joined forces to form Cometa Networks, an ambitious effort to start deploying wireless access points in 2003, with 20,000 WiFi hot spots set up throughout the country by this year. It was predicted that the number of hotspots could jump to 80,000 within four years.

Wireless local area networks (WLANs) are becoming ubiquitous and increasingly relied on as IEEE 802.11 products become successful in the market. In wireless home and office networks where voice, video, and audio will be delivered, quality of service (QoS) and multimedia support are critical, and are essential ingredients to offer customers video on demand, audio on demand, voice over IP, and high-speed Internet access.

Consumer electronics companies, such as Panasonic and Sharp, are looking to offer home wireless networking devices in which QoS is a critical element. In the near future, all consumer electronic devices at home, such as VCRs, TVs, and microwave ovens, can be equipped for and connected to QoS-enabled wireless networks. One of the most promising home networking candidates is IEEE 802.11 WLAN.

IEEE 802.11 medium access control (MAC) employs a mandatory contention-based channel access function called the distributed coordination function (DCF) and an optional centrally controlled channel access function called the point coordination function (PCF) [1]. The DCF adopts carrier sense multiple access with collision avoidance (CSMA/CA) and binary exponential backoff. It is treated as the wireless version of the most successful LAN, IEEE 802.3 (Ethernet), which adopts CSMA with collision detection (CSMA/CD) and binary exponential backoff. Both IEEE 802.11 DCF and IEEE 802.3 enable fast installation with minimal management and maintenance costs, and are very robust protocols for best effort service. However, the current DCF is unsuitable for multimedia applications with QoS requirements [2]. Even though the PCF can provide some limited QoS support, it is rarely implemented in today’s products due to its complexity and inefficiency for normal data transmission.

To support MAC-level QoS, the IEEE 802.11 Working Group is currently working on IEEE 802.11e [3–7], which is in the final stage. The emerging IEEE 802.11e standard provides QoS features and multimedia support to the existing 802.11b and 802.11a wireless standards, while maintaining full backward compatibility with these standards.
IEEE 802.11 MAC

IEEE 802.11 MAC employs a mandatory DCF and an optional PCF. These functions determine when a station (STA), operating within a basic service set (BSS) or independent BSS (IBSS), is permitted to transmit. There are two types of 802.11 networks: an infrastructure network (BSS) in which an access point (AP) is present, and an ad hoc network (IBSS) in which an AP is not present.

In a long run, time is always divided into repetition intervals called superframes. Each superframe starts with a beacon frame, and the remaining time is further divided into a contention-free period (CFP) and a contention period (CP). The DCF works during the CP, and the PCF works during the CFP. If the PCF is not active, a superframe will not include the CFP. However, the beacon frame is always sent no matter whether the PCF is active or not. The beacon frame is a management frame for synchronizations, power management, and delivering parameters. In a BSS, an AP sends beacon frames. In an IBSS, any mobile station that is configured to start an IBSS will begin sending beacon frames. As other mobile stations join that IBSS, each station that is a member of the IBSS is randomly chosen for the task of sending a beacon frame. Beacon frames are generated at regular intervals called target beacon transmission times.

The Distributed Coordination Function

The DCF defines a basic access mechanism and an optional request-to-send/clear-to-send (RTS/CTS) mechanism. In the DCF a station with a frame to transmit monitors the channel activities until an idle period equal to a distributed interframe space (DIFS) is detected. After sensing an idle DIFS, the station waits for a random backoff interval before transmitting. The backoff time counter is decremented in terms of slot time as long as the channel is sensed idle. The counter is stopped when a transmission is detected on the channel, and reactivated when the channel is sensed idle again for more than a DIFS. In this manner, stations deferred from channel access because their backoff time was larger than the backoff time of other stations are given higher priority when they resume their transmission attempt. The station transmits its frame when the backoff time reaches zero. At each transmission, the backoff time is uniformly chosen in the range (0, CW – 1) in terms of timeslots, where CW is the current backoff window size. At the very first transmission attempt, CW equals the initial backoff window size, CW_min. After each unsuccessful transmission, CW is doubled until a maximum backoff window size, CW_max, is reached. After the destination station successfully receives the frame, it transmits an acknowledgment frame (ACK) following a short interframe space (SIFS) time. If the transmitting station does not receive the ACK within a specified ACK timeout, or detects the transmission of a different frame on the channel, it reschedules the frame transmission according to the previous backoff rules.

The DCF provides a channel access mechanism with equal probabilities to all stations contending for the same wireless medium. If an AP is present, STAs are not allowed to transmit frames to other STAs that are not APs.

The above mechanism is called the basic access mechanism. In such a mechanism, the hidden node problem may happen: transmissions of a station cannot be detected using carrier sense by a second station, but interfere with transmission from the second station to a third station. To reduce the hidden station problem, an optional four-way data transmission mechanism, RTS/CTS, is also defined in DCF. In RTS/CTS, before transmitting a data frame, a short RTS frame is transmitted. The RTS frame also follows the backoff rules introduced above. If the RTS frame succeeds, the receiver station responds with a short CTS frame. Then a data frame and an ACK frame will follow. All four frames (RTS, CTS, data, ACK) are separated by an SIFS time. In other words, the short RTS and CTS frames reserve the channel for the data frame transmission that follows.

Since radio transmission has a range, the range (denoted RA) of the source’s RTS transmission and the range (denoted RB) of the destination’s CTS transmission are overlapped but not equal. Therefore, after transmission of the source’s RTS and the destination’s CTS, the channel is reserved for the data transmission that follows, and any station in either RA or RB will not transmit. A hidden station of the source, who is in RB but not in RA, will not interfere with the source data transmission since it hears the destination’s CTS transmission.

The Point Coordination Function

The PCF is an optional centrally controlled channel access function, which provides contention-free (CF) frame transfer. The PCF is designed to support time-bounded services, which can provide limited QoS. It logically sits on top of the DCF and performs polling, enabling polled stations to transmit without contending for the channel. It has higher priority than the DCF by adopting a shorter interframe space (IFS) called the point interframe space (PIFS); SIFS < PIFS < DIFS.

If the PCF is implemented, a CFP under the PCF and a CP under the DCF alternate over time. A CFP and a CP form a superframe. The AP where the point coordinator (PC) is normally located senses the medium idle for a PIFS interval, and then transmits a beacon frame to initiate a CFP (i.e., to initiate a superframe). After a SIFS time, the PC sends a poll frame to a station to ask to transmit a frame. The poll frame may or may not include data to that station. After receiving the poll frame from the PC, the station with a frame to transmit may choose to transmit a frame after a SIFS time. When the destination station receives the frame, an ACK is returned to the source station after a SIFS time. The PC waits a PIFS interval following the ACK frame before polling another station or terminating the CFP by transmitting a CF-End frame. If the PC receives no response from the polled station for a PIFS interval, the PC can poll the next station or terminate the CFP by transmitting a CF-End frame.
The PCF cannot provide good QoS support since it lacks an admission function to control channel access from stations. Therefore, when the traffic load is too high, all existing traffics may be degraded.

IEEE 802.11E

IEEE 802.11e provides a channel access function, the HCF, to support applications with QoS requirements. The HCF includes both contention-based and centrally controlled channel access. Contention-based access of the HCF is also referred to as the EDCF. In this article we refer to the centrally controlled channel access of the HCF as the controlled HCF.

QoS enhancements are available to QoS enhanced stations (QSTAs) that can be associated with a QoS enhanced access point (QAP) in a QoS BSS (QBSS) or in a QoS IBSS (QIBSS) without a QAP. Since a QSTA implements a superset of STA functionality, the QSTA may associate with a non-QoS AP in a non-QoS BSS to be backward compatible by providing non-QoS MAC data service. Note that the term non-QoS stands for the original MAC introduced in the previous section.

Similar to the original MAC, a CFP under the controlled-HCF and a CP under the EDCF alternate over time, and time is always divided into superframes. Each superframe starts with a beacon frame, and the remaining time is further divided into a CFP and a CP. The EDCF works during the CP, and the controlled HCF works during CFP. A new concept, transmission opportunity (TXOP), is introduced in IEEE 802.11e for both the EDCF and controlled HCF. A TXOP is a time period when a STA has the right to initiate transmissions onto the wireless medium. It is defined by a starting time and a maximum duration. A STA cannot transmit a frame that extends beyond a TXOP. If a frame is too large to be transmitted in a TXOP, it should be fragmented into smaller frames. A TXOP is used by both the EDCF and the controlled HCF, and can be obtained both when the medium is available under the EDCF rules and when the station receives a poll. Non-AP QSTAs may send TXOP requests during polled TXOPs or EDCF TXOPs using the QoS control field in the frame directed to the QAP, with the request duration or queue size indicated to the QAP.

**THE ENHANCED DCF**

The EDCF is based on differentiating priorities at which traffic is to be delivered and works with four access categories (ACs), which are virtual DCFs, as shown in Fig. 1, where each AC achieves a differentiated channel access. This differentiation is achieved through varying the amount of time a STA will sense the channel to be idle and the length of the contention window during backoff. The EDCF supports eight different priorities, which are further mapped into four ACs, shown in Table 1. ACs are achieved by differentiating the arbitration interframe space (AIFS), initial window size, and maximum window size. For AC $i$ ($i = 0, ..., 3$), the initial backoff window size is $CW_{\min}[i]$, the maximum backoff window size is $CW_{\max}[i]$, and the AIFS is $AIFS[i]$. For AC $i$ ($i = 0, ..., 3$), the backoff mechanism is similar to the original MAC: after each unsuccessful transmission, $CW[i]$ is doubled until a maximum backoff window size value $CW_{\max}[i]$ is reached. For $0 \leq i < j \leq 3$, we have $CW_{\min}[i] \geq CW_{\min}[j]$, $CW_{\max}[i] \geq CW_{\max}[j]$, and $AIFS[i] \geq AIFS[j]$, and at least one of the above inequalities must be “not equal to.” In other words, the EDCF employs $AIFS[i]$, $CW_{\min}[i]$, and $CW_{\max}[i]$ (all for $i = 0, ..., 3$) instead of DIFS, $CW_{\min}$, and $CW_{\max}$, respectively. If one AC has a smaller AIFS or $CW_{\min}$ or $CW_{\max}$, the AC’s traffic has a better chance of accessing the wireless medium earlier.

Figure 2 shows the EDCF timing diagram, where three ACs are shown: $i$, $j$, and $k$. Figure 1 shows four transmission queues implemented in a station, and each queue supports one AC, behaving roughly as a single DCF entity in the original IEEE 802.11 MAC. It is assumed that a payload from a higher layer is labeled with a priority value and enqueued into the corresponding

---

1. In an earlier version of the IEEE 802.11e draft, up to eight queues were implemented instead of four before the concept of ACs was introduced.

<table>
<thead>
<tr>
<th>Priority</th>
<th>AC</th>
<th>Designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>Best effort</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Best effort</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Best effort</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Video probe</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>Video</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>Voice</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>Voice</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>Voice</td>
</tr>
</tbody>
</table>

**Table 1. Priority to access category mapping.**
queue according to the mapping in Table 1. Each queue acts as an independent MAC entity and performs the same DCF function, with a different interframe space \( AIFS[i] \), a different initial window size \( CW_{\text{min}}[i] \), and a different maximum window size \( CW_{\text{max}}[i] \). Each queue has its own backoff counter \( BO[i] \), which acts independently in the same way as the original DCF backoff counter. If there is more than one queue finishing the backoff at the same time, it is called an internal collision, which is resolved as follows. The highest AC frame is chosen to transmit by the virtual collision handler. Other lower AC frames whose backoff counters also reach zero will increase their backoff counters with \( CW_{\text{min}}[i] \) \( (i = 0, ..., 3) \) accordingly. In other words, other lower AC frames whose backoff counters also reach zero do not treat this internal collision as a real collision after adjusting the backoff counters. Furthermore, we have \( AIFS[i] \geq PIFS \).

The values of \( AIFS[i] \) \( (i = 0, ..., 3) \), \( CW_{\text{min}}[i] \) \( (i = 0, ..., 3) \), and \( CW_{\text{max}}[i] \) \( (i = 0, ..., 3) \) are referred to as the EDCF parameters, announced by the QAP via periodically transmitted beacon frames. The QAP can also adaptively adjust these EDCF parameters based on the network traffic conditions.

**DISTRIBUTED ADMISSION CONTROL FOR EDCF**

The QoS parameter set element (QPSE) provides information needed by QSTAs for proper operation of the QoS facility during the CP. The QPSE includes \( CW_{\text{min}}[i] \), \( CW_{\text{max}}[i] \), \( AIFS[i] \), \( TXOPLimit[i] \), \( TXOPBudget[i] \), \( Load[i] \), and \( SurplusFactor[i] \) for \( i = 0, ..., 3 \). For the AC \( i \), \( TXOPLimit[i] \) specifies the time limit on TXOPs; \( TXOPBudget[i] \) specifies the additional amount of time available during the next beacon interval; \( Load[i] \) specifies the amount of time used during the previous beacon interval; and \( SurplusFactor[i] \) represents the ratio of over-the-air bandwidth reserved to the bandwidth of the transported frames required for successful transmission. The QPSE is calculated by the QAP for each beacon interval and embedded into the next beacon frame transmitted to each QSTA. When the transmission budget for an AC is depleted, new QSTAs cannot gain transmission time, while existing QSTAs cannot increase the transmission time they are already using. This mechanism protects existing flows.

The QAP shall measure the amount of time occupied by transmissions for each AC during the beacon period, including associated SIFS and ACK times if applicable. The QAP shall maintain a set of counters \( TXTime[i] \), which shall be set to zero immediately following transmission of a beacon. For each data frame transmission (either uplink or downlink), the QAP shall add to the \( TXTime \) corresponding to the AC that frame a time equal to the frame transmission time and all overhead involved (SIFS time, MAC and PHY overhead, ACK, etc.). The QAP shall transmit in each beacon the \( TXOPBudget \) and \( SurplusFactor \) for each AC, contained in QPSE. The QAP will set \( TXOPBudget[i] \) to be

\[
TXOPBudget[i] = \max(ATL[i] - TXTime[i]*SurplusFactor[i], 0).
\]

The variable \( ATL[i] \) is the maximum amount of time that may be spent on transmissions of AC \( i \) per beacon interval.

Each QSTA has to maintain the following variables for each AC: \( TxCounter[i] \), \( TxUsed[i] \), \( TxLimit[i] \), \( TxRemainder[i] \), and \( TxMemory[i] \). The variable \( TxUsed[i] \) counts the amount of time occupied on air by transmissions, successful or otherwise, from this station for this AC, including associated SIFS and ACK times if applicable. \( TxCounter[i] \) counts successful transmissions. The STA shall not transmit a data frame if doing so would result in the value in \( Txused[i] \) exceeding the value of \( TxLimit[i] \). If the QSTA is prevented from sending a frame for this reason, it may carry over the partial frame time remainder to the next beacon period, by storing the remainder in \( TxRemainder[i] \); \( TxRemainder[i] = TxLimit[i] - TxUsed[i] \); otherwise, \( TxRemainder[i] = 0 \). At each target beacon transmission time, the \( TxMemory \), \( TxLimit \), and \( TxCounter \) state variables are updated according to the following procedure:

- **If** \( TXOPBudget[i] = 0 \) 
  - \( -TxMemory[i] \) shall be set to zero for new QSTAs that started transmission with this AC in the last beacon interval; all other QSTAs \( TxMemory[i] \) remains unchanged
- **If** \( TXOPBudget[i] > 0 \)
DLP allows QSTAs to transmit frames directly to another QSTA by setting up such data transfer when a QAP is present. The need for this protocol is motivated by the fact that the intended recipient may be in Power Save Mode, in which case it can only be woken up by the QAP.

\[ TxMemory[i] = f \cdot TxMemory[i] + (1 - f) \cdot (TxCounter[i] \cdot SurplusFactor[i] + TXOPBudget[i]) \]
\[ TxCounter[i] = 0 \]
\[ TxLimit[i] = TxMemory[i] + TxRemainder[i] \]
where \( f \) is the damping factor, which does not affect the entrance of a new flow into the system when enough budget is available, because the decreased TXOPBudget is offset by an increased TxCounter instantaneously, so TxMemory does not change a lot. The damping does affect TxMemory when a new flow starts up in another QSTA. In that case, the decreased TXOPBudget is not offset by an increased TxCounter; consequently, the TxMemory converges to the lower target value.

The TXOPBudget used in this calculation shall be the budget most recently obtained from the QAP. The TxCounter value shall be the value of the beacon period before the period that just ended. Taking the earlier value accounts for the delay that occurs between the moment at which the QAP determines the TXOPBudget and the point at which this budget is used in the above calculations.

The value TxCounter + TXOPBudget is the target to which TxMemory converges. TxLimit is equal to TxMemory plus a possible capped remainder. TxMemory “memorizes” the amount of resources the QSTA has been able to spend in a specific AC. Once the budget is depleted (i.e., TXOPBudget hovers around 0), TxMemory converges to TxCounter, which is the lower limit. This ensures that the QSTA can continue consuming the same amount of resources in following beacon periods. Damping allows for some amount of fluctuation to occur, but TxMemory cannot grow any further in the saturated state. This prevents new flows from entering the specific AC when it is saturated. A suitable initial value for this variable could be between 0 and TXOPBudget[i]/SurplusFactor[i].

QSTAs shall not increase their TxLimit[i] if they did not transmit traffic with the AC in the previous beacon interval.

**Direct Link Protocol**

Direct Link Protocol (DLP) allows QSTAs to transmit frames directly to other QSTAs by setting up such data transfer when a QAP is present. The need for this protocol is motivated by the fact that the intended recipient may be in power save mode, in which case it can only be woken up by the QAP. DLP allows the sender and receiver to exchange rate set and other information. Furthermore, DLP messages can be used to attach security information elements.

This protocol prohibits STAs going into power save for the active duration of the direct stream. DLP does not apply in an ad hoc network, where frames are always sent directly from one QSTA to another. A direct link can be built by the following sequences:

- **QSTA-1** that has data to send invokes DLP and sends a DLP-request frame to the QAP. This request contains the rate set, and (extended) capabilities of QSTA-1, as well as the MAC addresses of QSTA-1 and QSTA-2.
- **QSTA-2** that receives the DLP-request frame sends a DLP-response frame to the QAP. The QAP forwards the DLP-response frame to the QSTA-1. This mechanism is expected to be used for applications such as voice and video, which may need periodic service from the HC. Non-QoS STAs may also associate in a QBSS. All frames that are sent to QSTA-1, after which the direct link becomes active and frames can be sent from QSTA-1 to QSTA-2 and from QSTA-2 to QSTA-1.

When the direct link is active, QSTA-1 may use DLP probes to gauge the quality of the link between QSTA-1 and QSTA-2. The direct link becomes inactive when no frames have been exchanged as part of the direct link for the duration of a DLPIdleTimeout. After the timeout, frames with destination QSTA-2 shall be sent via the QAP.

**The Controlled HCF**

The controlled HCF, the centrally controlled channel access function, allows reservation of transmission opportunities (TXOPs) with a hybrid coordinator (HC), a type of PC handling rules defined by the HCF. The HC, in conjunction with a QAP, performs bandwidth management including allocation of TXOPs to QSTAs. Similar to a PC, an HC can transmit the beacon frame to initiate CFP when it senses the medium idle for a PIFS interval, and terminate the CFP by transmitting a CF-End frame when it senses the medium idle for a PIFS interval.

A non-AP QSTA, based on its requirements, requests the HC for TXOPs for both its own transmissions and transmissions from the HC to itself. Non-AP QSTAs may send TXOP requests using the QoS control field in the frame directed to the QAP, with the request duration or queue size indicated to the QAP. The HC, based on an admission control policy introduced in the next subsection, either accepts or rejects the request. If the request is accepted, it schedules TXOPs for the non-AP QSTA. For transmissions for the STA, HC polls a non-AP QSTA based on the parameters supplied by the non-AP QSTA at the time of its request. For transmissions to the non-AP QSTA, the HC queues the frames and delivers them periodically, again based on the parameters supplied by the non-AP QSTA. This mechanism is expected to be used for applications such as voice and video, which may need periodic service from the HC. Non-QoS STAs may also associate in a OBSS. All frames that are sent to non-QoS STAs by an AP conform to the original MAC introduced earlier.

**Admission Control and Scheduling for the Controlled HCF**

When the HC provides controlled channel access to non-AP QSTAs, it is responsible for granting or denying polling service based on the admitted traffic streams. The behavior of the scheduler is as follows:

- The scheduler shall be implemented such that, under the controlled HCF, all stations with admitted traffic streams are offered TXOPs that satisfy the service schedule.
- Specifically, if a traffic stream is admitted by the HC, the scheduler shall send polls any-
The HC aggregates admitted traffic streams for a single QSTA and establishes a service schedule for the QSTA. During any time interval $[t_1, t_2 - D]$, the cumulative TXOP duration shall be greater than the total time required to transmit all frames belonging to all the admitted streams, each arriving at the mean data rate for the stream, over the period $[t_1, t_2 - D]$. The parameter $D$ is set to the specified maximum service interval. If the maximum service interval is not specified, $D$ is set to the delay bound.

A minimum set of traffic stream parameters shall be specified during the negotiation. The specification of a minimum set of parameters is required so that the scheduler can determine a schedule for an admitted stream. These parameters are mean data rate, nominal MSDU size, and at least one of maximum service interval and delay bound.

Note that IEEE 802.11e specifies some minimum mandatory requirements for admission control and scheduling for the controlled HCF as discussed above, and allows vendors to design their own admission control and scheduler based on some guidelines. An informative example is provided for vendors’ consideration in IEEE 802.11e as follows.

An example of a traffic stream is defined by a set of parameters such as mean data rate, nominal frame size, and maximum service interval or delay bound. The maximum service interval is the maximum interval between the start of two successive QoS CF polls. All of these criteria affect the admissibility of a given traffic stream, and any new traffic stream may be rejected. If both maximum service interval and delay bound are specified, uses the maximum service interval for calculation of the schedule. The schedule for an admitted stream is calculated in two steps:

- Calculation of the scheduled service interval (SI)
- Calculation of TXOP duration for a given SI for the stream

First, calculate the minimum of all maximum SI for all admitted streams. Let this minimum be $m$. Second, the scheduler chooses a number lower than $m$ that is a submultiple of the beacon interval. This value is the scheduled SI for all QSTAs with admitted streams.

When a new stream requests admission, the admission control process is done in three steps. First, calculate the number of frames that arrive at the mean data rate during the scheduled SI. Second, calculate the TXOP duration that needs to be allocated for the stream. Finally, the admission control unit (ACU) determines that the stream can be admitted when the following inequality is satisfied, where $k$ is the number of existing streams and $k + 1$ is used as the index for the newly arriving stream. $T$ indicates the beacon interval and $T_{CP}$ is the time used for EDCF traffic.

$$\frac{\text{TXOP}_{k+1}}{\text{SI}} + \sum_{j=1}^{k} \frac{\text{TXOP}_j}{\text{SI}} \leq \frac{T - T_{CP}}{T}. \quad (2)$$

All admitted streams have guaranteed access to the channel. For the calculation of the TXOP duration for an admitted stream, the simple scheduler uses the following parameters: mean data rate ($r$) and nominal MSDU size ($L$) from the negotiated traffic stream, the scheduled SI ($\text{SI}$) calculated above, physical transmission rate ($R$), size of maximum frame (2304 bytes, $M$), and overheads in time units ($O$). The physical transmission rate is the minimum PHY rate negotiated in the traffic stream. The overheads in time includes interframe spaces, ACks, and CF polls.

For the calculation of the TXOP duration for an admitted stream, the simple scheduler uses the following parameters. First, the scheduler calculates the number of frames that arrived at the mean data rate during the SI: $N_i$; then the scheduler calculates the TXOP duration as the maximum of:

1. Time to transmit $N_i$ frames at $R_i$
2. Time to transmit one maximum size MSDU at $R_i$ (plus overheads):

   $$N_i = \left[ \frac{\text{SI} \cdot r_i}{L_i} \right]$$

   $$\text{TXOP}_i = \max \left( \frac{L_i \cdot N_i}{R_i} + O, \frac{M}{R_i} + O \right)$$

An example of the scheduling is shown in Fig. 3. Stream from QSTA $i$ is admitted in Fig. 3a. The beacon interval is 100 ms, and the maximum SI for the stream is 60 ms. The scheduler calculates a scheduled SI equal to 50 ms using the steps explained above. The same process is repeated continuously while the maximum SI for the admitted stream is smaller than current SI. If a new stream is admitted with a maximum SI smaller than the current SI, the scheduler needs to change the current SI to a smaller number than the maximum SI of the newly admitted stream. Therefore, the TXOP duration for the current admitted streams also needs to be recalculated with the new SI. If a stream is dropped (Fig. 3c), the scheduler might use the time available to resume contention. The scheduler might also choose to move the TXOPs for the QSTAs following the QSTA dropped to use the unused time. However, this last option might require the announcement of a new schedule to all QSTAs. Fig. 3c shows the situation when a stream for QSTA $j$ is removed.

**The Group Acknowledgment Mechanism**

Since the wireless medium is error-prone, transmitted frames can easily be corrupted even without collisions. In the IEEE 802.11 MAC protocol, each frame is acknowledged. This approach is very natural and robust, but introduces a great deal of overhead. In order to reduce the acknowledgment overhead, a new mechanism, called group acknowledgment (GA), is introduced in IEEE 802.11e.

The GA mechanism allows a group of frames to be transmitted before any acknowledgment. After sending a burst of frames, the sender sends a group acknowledgment request (GroupAckReq) frame, and the receiver must respond by sending the group acknowledgment...
The EDCF provides a priority scheme by differentiating the inter-frame space, the initial window size, and the maximum window size. Such a prioritized QoS scheme is simple to implement and is similar to the DiffServ model.

The sender should first win a TXOP using a channel access mechanism before starting a burst. The burst length is limited, and the amount of state that must be kept by the receiver of the receiving frames is bounded. The GroupAck frame has an Ack bitmap field that can acknowledge the burst of data frames, where each bit acknowledges one potential frame. If the GroupAck indicates that a frame was not received correctly, the sender shall retry that frame subject to its appropriate retry limit. Retransmitted burst data frames shall preserve their original relative order.

The receiver shall maintain a burst acknowledgment record consisting of a transmitter address and a 32-octet bitmap of received frame sequence numbers. These hold the acknowledgment state of the burst data received from that sender.

**Preliminary Results**

We evaluate the EDCF priority scheme in this section via simulations. The simulation models had been developed based on the IEEE 802.11e draft [3], IEEE 802.11a standard [9], and OPNET WLAN simulation model version 8.0A. In our simulations, distributed admission control is not considered. A more complete study of priority schemes can be found in [7].

We simulate three priority classes in Fig. 5. Both the data rate and control rate are 6 Mb/s. The frame size is fixed at 1024 bytes. Each station/queue always has frames ready to send to achieve saturation performance. The EDCF parameters are:

- \( AIFS[0] = PIFS + 1 \mu s; AIFS[1] = PIFS + 5 \mu s; AIFS[2] = DIFS \); therefore, \( AIFS[0] < AIFS[1] < AIFS[2] \); where \( PIFS = SIFS + SLOT = 25 \mu s \) and \( DIFS = SIFS + 2*SLOT = 34 \mu s \).
- \( CW_{\min}[0] = 16; CW_{\min}[1] = 24; CW_{\min}[2] = 32 \); therefore, \( CW_{\min}[0] < CW_{\min}[1] < CW_{\min}[2] \).

Figure 5 shows normalized saturation throughputs for three priority classes over the number of active queues. The number of active queues in the figure stands for the number of active queues for each class. As the number of active queues increases, throughputs for all classes decrease. As illustrated in the figure, class 0 has a much better throughput than class 1, and class 1 has a much better throughput than class 2. Therefore, the EDCF priority scheme is quite effective.
CONCLUSIONS AND FUTURE WORK

This article introduces the emerging IEEE 802.11e MAC, which employs a contention-based channel access EDCF and a centrally controlled channel access HCF. The controlled HCF is based on a polling mechanism with some enhanced QoS-specific mechanisms and frame subtypes to allow QoS data transfers during CFP. The EDCF provides a priority scheme by differentiating the interframe space, initial window size, and maximum window size. Such a prioritized QoS scheme is simple to implement and is similar to the differentiated services model. Distributed admission control under the EDCF, centrally controlled admission control and a simple scheduler under the controlled HCF, the direct link protocol, and the group acknowledgment mechanism are also introduced. Our preliminary results show that IEEE 802.11e provides very good service differentiation.

REFERENCES


Figure 5. Normalized saturation throughputs of three classes.

BIographies

Yang Xiao [M] (yangxiao@ieee.org) is an assistant professor in the Computer Science Division at the University of Memphis, Tennessee. Before joining this school in August 2002, he worked at Micro Linear as a MAC architect involving IEEE 802.11 standard enhancement work. He is a voting member of the IEEE 802.11 Working Group, and holds a Ph.D. degree in computer science and engineering from Wright State University, Dayton, Ohio. His current research interests include WLANs, WPANs, and mobile cellular networks.