Introduction to IEEE 802.11 standards

(Z. MAMMERI – UPS)

I.1. QoS in Wireless networks

IEEE 802.11 is the most popular standard for wireless LANs but lacks QoS provisioning capabilities. Recently, a new standard IEEE 802.11e has been proposed for providing Quality of Service in the IEEE 802.11 LANs. The IEEE 802.11e operates in two modes: a distributed mode and a centralized one. For providing QoS guarantees, the centralized approach is the most promising. The IEEE 802.11e standard proposes a Reference Scheduler and an Admission Control Unit (ACU) for admission and management of traffic flows in the centralized approach. Flows register with the scheduler and then, based on their QoS requirements, the Reference Scheduler works in a simple round robin manner to allocate transmission times to stations containing the flows. The reference scheduler works fine for constant bit-rate traffic (CBR), but is inefficient for Variable bit-rate traffic (VBR) because it allocates times to stations based on the average traffic specs of the flows. Numerous multimedia applications such as quality controlled MPEG4 and video conferencing have VBR characteristics and it is thus mandatory that mechanisms must be present to ensure QoS for these applications.

In this section, we present IEEE 802.11 (i.e. WiFi) and its limitations for providing QoS and then we present the IEEE 802.11e standard and its provisions for QoS. It should be noticed that much work has been done for providing QoS in wireless and ad hoc networks ([MAN02], [NI04], [NI05], [RAM05]), but will not be discussed in this document. We focus only on IEEE 802.11 because we believe that it will be the technology the most used for providing QoS.

I.1.1. IEEE 802.11 Standard (WiFi)

The IEEE 802.11 standard is the dominant standard for Wireless Local Area Networks. In 1997 IEEE adopted it as the first WLAN standard. The IEEE 802.11 standard is relating to the Medium Access Control (MAC) layer and the Physical (PHY) Layer. Figure 1 shows the MAC layer and the physical layer standards.
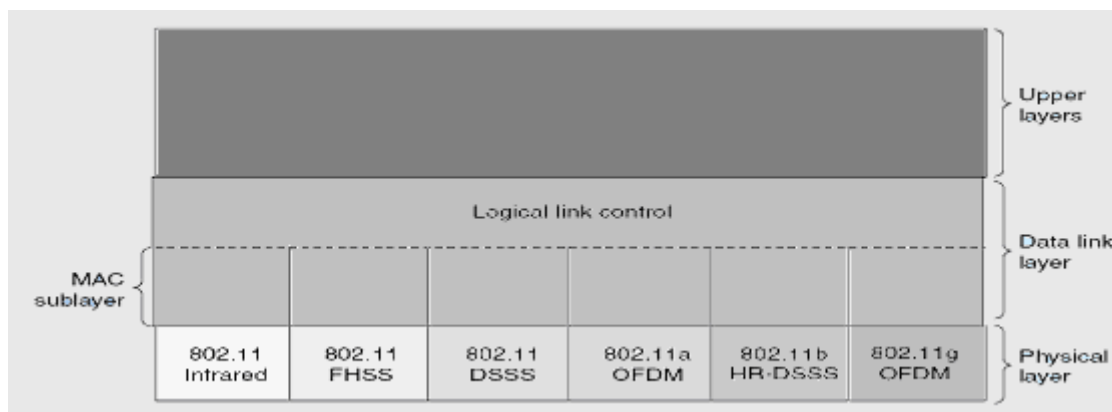

Figure 1. MAC sub-layer and the related physical layer standards for IEEE 802.11

a) Physical Layer

There is only one MAC layer, but different standards have been proposed for the Physical layer. The standard defines several different modulation methods: Infrared, Direct sequence spread spectrum (DSSS), Frequency-Hopping spread spectrum (FHSS), Orthogonal Frequency Division Multiplexing

(OFDM) and also defines three different Physical layer technologies: IEEE 802.11a, IEEE 802.11b and IEEE 802.11g.

IEEE 802.11a operates in the 5 GHz band. Theoretically, it provides a data rate of 54 Mbps, but more realistically it achieves throughput somewhere between 20 Mbps to 25 Mbps in normal traffic conditions. IEEE 802.11b operates in the 2.4 GHz band and has a maximum theoretical data rate of up to 11 Mbps with an average throughput of somewhere between 4 Mbps to 6Mbps. IEEE 802.11g offers delivers rates from 6 Mbps to 54 Mbps. It uses the OFDM modulation.

b) Basic Architecture

The IEEE 802.11 architecture is composed of several basic components. Some of the important components and functions are the following (Figure 2):

- *Station (STA)*: Any device that contains an IEEE 802.11 conformant medium access control (MAC) and physical layer (PHY) interface to the wireless medium (WM).

- *Access Point (AP)*: Any entity that has station functionality and provides access to the distribution services via the wireless medium (WM) for associated stations.

- *Basic Service Set (BSS)*: The *Basic Service Set* (BSS) is the basic building block of a IEEE 802.11 wireless LAN. The BSS consists of a group of any number of stations.

- *Basic Service Area (BSA)*: The conceptual area within which, the members of the Basic Service Set may communicate.
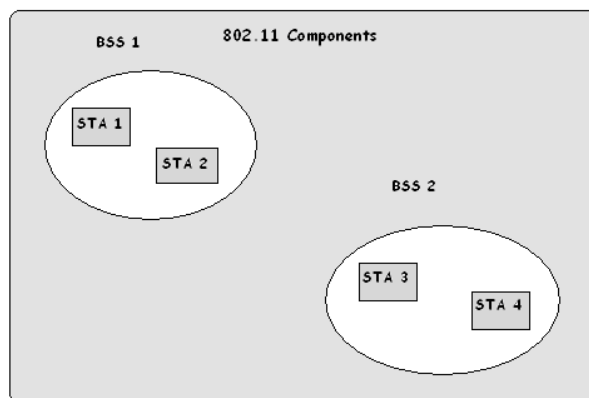

Figure 2. Basic Service Set (BSS) of IEEE 802.11 WLAN

c) Distribution System (DS)

The *distribution system* (DS) is the system used to interconnect a set of BSSs through the Access points. An Access Point in one BSS may communicate with another access point to exchange frames for stations in their respective BSSs.

d) Interframe space (IFS)

In IEEE 802.11 spacing is used to separate frames. The length of interframe space determines when the channel can be accessed. Thus, the interframe spacing is used to set prioritized access to the channel. There are four types of interframe spaces, *Short Interframe space* (SIFS), *PCF Interframe space* (PIFS), *DCF Interframe space* (DIFS) and *Extended Interframe space* (EIFS).

Figure 3 shows the relationship between the different Interframe spaces. As we can see, the relationship between these interframe spaces is: SIFS < PIFS < DIFS < EIFS
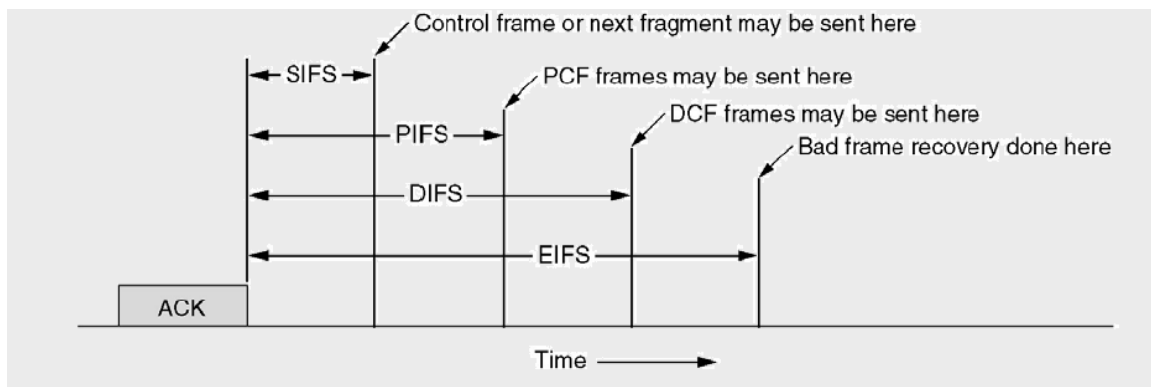
Figure 3. Different Interframe Spaces used in the IEEE 802.11

- *SIFS* is the shortest Interframe space and is used for separating packets within a single transmission and for the time to be waited before acknowledgement from the receiving station can be sent.

- *PIFS* is used by the Access Point (AP) to get prioritized access to the channel before any other stations can claim access.

- *DIFS* is used by stations to gain access to the medium during the contention phase i.e. when the Access Point does not govern the access to the channel. The wireless stations must sense the channel idle for DIFS before trying to access the channel.

- *EIFS* is extended interframe space used whenever there is an error in transmission. The station that transmitted the frame that was not correctly received (the station concludes this when it does not receive an ACK) must wait for EIFS before trying to transmit again.

e) Hidden Node Problem and RTS/CLS mode

The hidden node problem is one of the most common problems in Wireless Networks. This problem takes place in scenarios where there are nodes that are outside the range of each other but within the range of a common receiver. If both (or more) of these stations transmit to the common station simultaneously, then there is a collision.

The IEEE 802.11 standard provides a mechanism to solve this problem called "*Request-to-Send*" and "*Clear-to-Send*" (RTS/CLS) scheme. The station wishing to send data first sends a Request-to-Send (RTS) frame to the intended destination station. Upon the reception of this frame, the receiving station sends a Clear-to-Send (CTS) frame. Both the RTS and CTS frames contain the duration of the communication that is going to take place. This essentially solves the hidden-node problem, but there is an overhead associated with the extra messages that have to be exchanged for this scheme.

f) IEEE 802.11 Network Topologies

The network topology of the IEEE 802.11 WLAN has the basic building block Basic Service Set (BSS).There are two types of LAN topologies for the Wireless LANs, an *Independent Basic Service Set* (IBSS) and an *Infrastructure Basic Service Set* (Figure 4).

The independent BSS (IBSS) is the most basic type of IEEE 802.11 LAN. Stations are connected to each other through the wireless medium on a peer-to-peer basis. However, all the stations in the IBSS may not be able to communicate with each other. This mode of operation is possible when IEEE 802.11 stations are able to communicate directly. Because this type of IEEE 802.11 LAN is often formed

without pre-planning, for only as long as the LAN is needed, this type of operation is often referred to as an *ad hoc network*.

An Infrastructure Basic Service Set is a BSS with the central component of *Access Point* (AP) which performs the relay function. All stations in the BSS communicate through the AP. Some advantages are that it is simpler since stations no longer need to maintain neighbour relationships, it can improve the range of the BSA and we may utilize some power saving mechanisms for the stations. One station (STA) can associate with only one AP at one time and it is done through the process of *association*. The AP may be attached to a wired or wireless Ethernet and hence may provide connection to the Internet for the nodes.
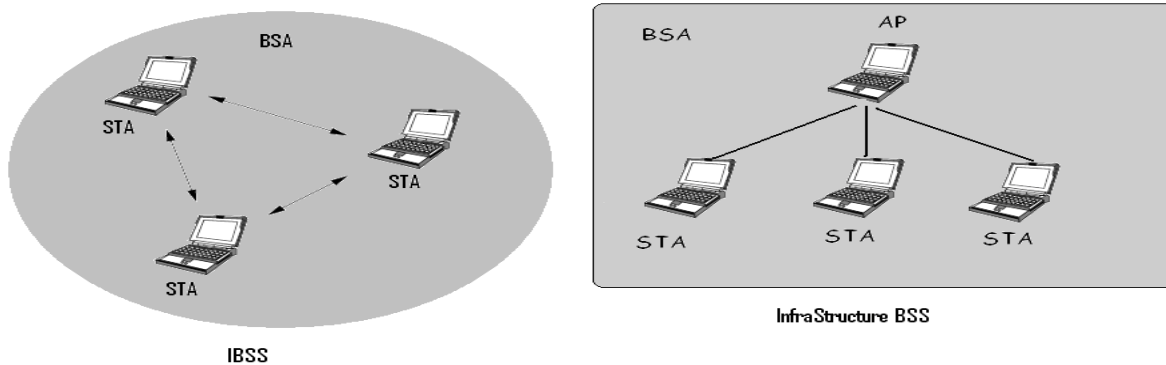
Figure 4. Independent Basic service Set and Infrastructure Basic service Set

g) IEEE 802.11 Medium Access Modes

IEEE 802.11 standard offers two different modes of operation: a mandatory Distributed Coordination Function (DCF) and an optional Point Coordination Function (PCF).

The basic IEEE 802.11 MAC protocol for accessing the cannel is the Distributed Coordination Function. Mechanisms for accessing the channel are independent in the sense that there is not central mechanism and stations try on their own to access the channel. It uses the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) i.e. listen before talk. Before transmitting each wireless station senses the medium and if the medium is idle for DCF interframe space (DIFS) the frame is transmitted.

Figure 5 shows the case of successful transmission from Source to Destination without using retransmission and backoff procedure.
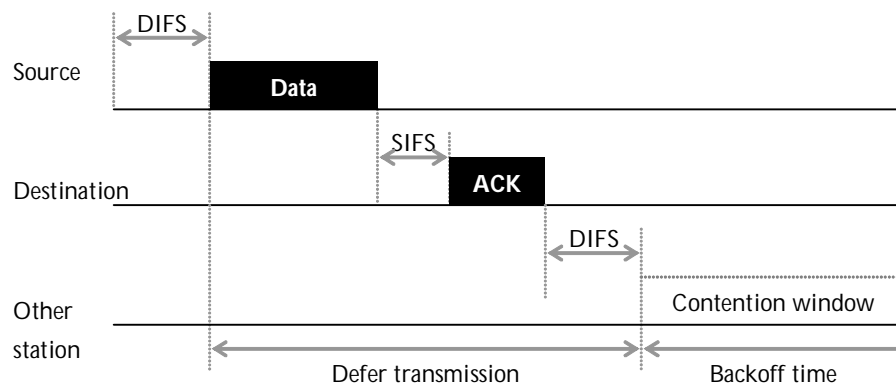
Figure 5. Nominal functioning of CSMA/CA without retransmission.

If however, the medium was busy, then the station chooses a random back-off timer which is measured in terms of time slots. This random back-off is selected on basis of a Contention Window (CW) and is chosen in the interval [0, CW].

After the medium has been idle for at least a DIFS, the back-off timer is decremented by one for each time slot that the medium remains idle. When the timer becomes zero, the frame is transmitted. If however, during the decrementing of the back-off timer, the medium becomes busy, the timer is frozen during that time and resumed after the channel is sensed idle again for more than DIFS.

There are collisions when more than one wireless stations start to transmit simultaneously. Hence, for being ensured of the successful transmission, upon the reception of a frame, the receiver sends an Acknowledgement (ACK) frame to the sender. This ACK frame is sent after the SIFS. If however, there is no ACK frame received, then the station assumes that there has been a collision. It reschedules the frame and enters the back-off procedure again. To reduce the probability of collisions, the station doubles the contention window for each retransmission attempt. The Collision window can be expanded up to a maximum limit of CWmax. After a successful transmission, the contention window is reset to CWmin. Even after a successful transmission, before transmitting the next frame, a backoff is done, which is called "post-backoff".

Figure 6 shows a scenario where five stations are competing to access the channel. First, we assume that station *A* is transmitting its data frame. During *A*'s transmission, new frames arrive at MAC sublayer of stations *B*, *C* and, D. Since the channel is idle, stations *B*, *C,* and *D* must differ their transmission till the end of *A*'s transmission. When the channel is sensed idle, stations *B, C,* and *D* randomly determine their backoff values before attempting access the channel. Station *B* (with the lowest backoff value) is the next station to win the competition. Stations *C* and *D* must wait (and their backoff timers are frozen). At the end of *B*'s transmission, station *D* is the next to win the channel access. Station *C* should still wait. However during *D*'s transmission a new frame arrives at station *E* MAC sublayer. At the end of *D*'s transmission, station *E* determines its random backoff value, which is less than the remaining backoff of station *C*. Thus station *E* wins the channel access before station *C*. At the end of *E*'s transmission, station *C* is alone to compete, thus it can send its data.
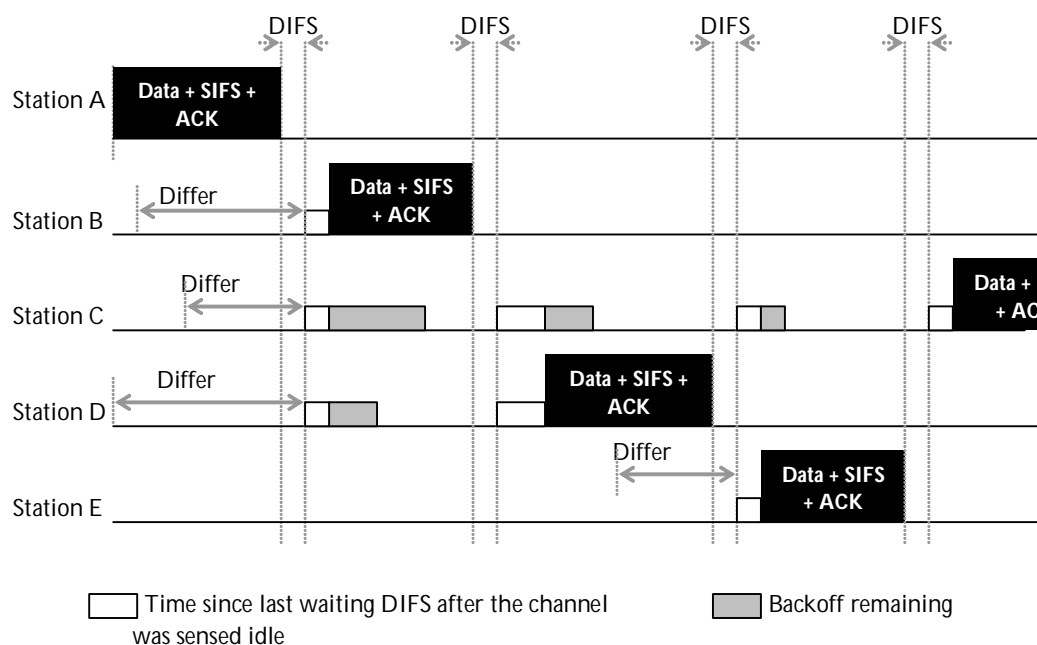

Figure 6. Example of competition under CSMA/CA.

The IEEE 802.11 standard also defines the *Point Coordination Function* (PCF) which uses a central *Point Coordinator* (PC) which manages access to the wireless medium and polls stations during a *Contention Free Period* (CFP). During CFP, access to the medium is governed and assigned to stations

by the Point Coordinator, but during the *Contention Period* (CP), there is no central authority and all the stations contend for access to the medium. The contention-based period and the contention-free period alternate and together they form the *Super-Frame*. The PC shall determine the CFP-Rate and this value is communicated to other STAs in the BSS in the Beacon frames.

PCF uses *the PCF Inter Frame Space* (PIFS) for accessing the medium, which is shorter than the DIFS and thus has a higher priority. The PC periodically generates a Beacon Frame which is used for timing synchronization and other things. The PC maintains a list for all the stations that it has to poll and the PC polls each station according to the list. The poll frame that tells a station to transmit may be accompanied by some data that the receiving station then acknowledges. At the end of Contention-Free period, a frame called CF-End is transmitted to signal the end of Contention-Free period. The CFP is started by sending a beacon frame during the contention period and since this is contention based access, the CFP may be delayed and hence shortened in its duration. Figure 7 shows the CP and the CFP and a shortened CFP.
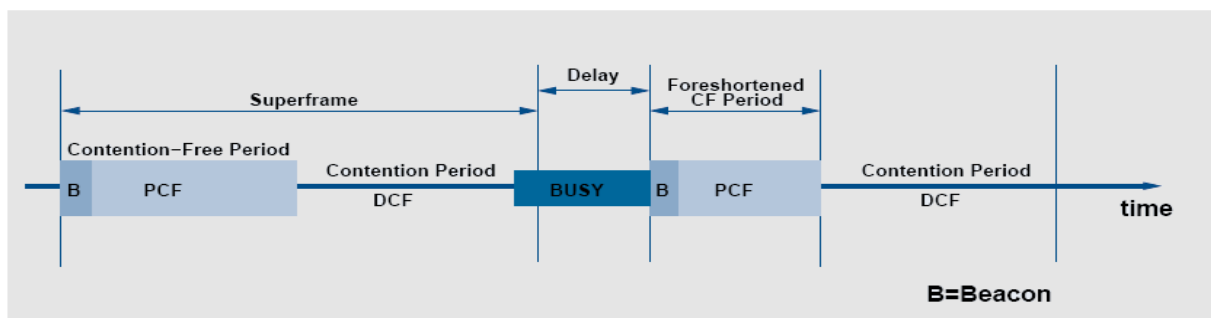

Figure 7. Alternating CP and CFP

I.1.2. IEEE 802.11e standard for QoS

The IEEE 802.11 standard, although being widely used, is not capable of providing QoS guarantees to flows [IEE05]. Since QoS is an increasing need in view of the increasing number of multimedia and other applications which depend on such a mechanism, efforts were made to introduce a new standard. Therefore, the IEEE 802.11e standard was introduced to address the shortcomings of IEEE 802.11 for providing QoS. The IEEE 802.11e standard proposes some major changes to the IEEE 802.11 standard to incorporate the QoS provisioning services. IEEE 802.11e introduces a new MAC layer function, called the *Hybrid Coordination Function* (HCF). The HCF introduces some components and techniques that were missing in the IEEE 802.11 standard. Some of the most important new terms and concepts that IEEE 802.11e has introduced are the following:

- *QoS Station (QSTA)*: A Station that implements the QoS facility.
- *QoS BSS (QBSS)*: A BSS that implements the QoS facility.
- *QoS Access Point (QAP)*: An Access Point that supports the QoS of the standard.
- *Transmission Opportunity (TXOP)*: Time duration during which a QSTA can send a burst of data.
- *Hybrid Coordinator (HC)*: A centralized controller for all other stations in the QBSS.
- *Controlled Access Phase (CAP)*: Time during which the HC maintains control of the medium.

The HCF operates in two modes: the contention-based *Enhanced Distributed Coordination Function* (EDCF) and the contention-free *HCF Controlled Channel Access* (HCCA). The HCF function of IEEE 802.11 works on top of the DCF function. Figure 8 illustrates the relationship between the DCF and PCF of IEEE 802.11 and the new EDCF and HCCA functions of IEEE 802.11e:
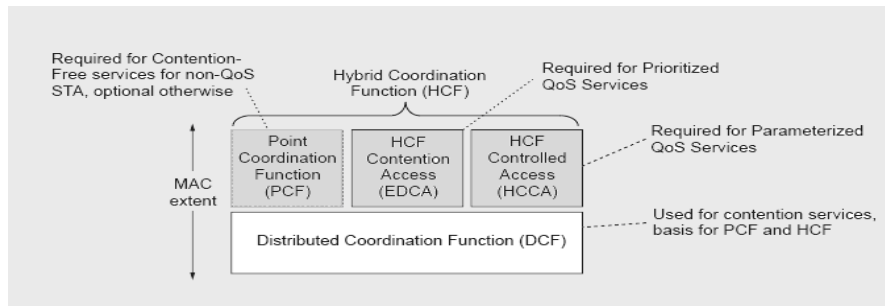
Figure 8. Relationship between DCF, PCF, EDCF and HCCA

a) Enhanced Distributed Coordination Function (EDCF)

The EDCF is the contention based part of the HCF. It works on the same principles as the DCF of IEEE 802.11, but is aimed to improve the contention-based access to channel to provide priorities. The EDCA introduces at MAC layer four *Access Categories* (ACs) supporting eight *User Priorities* (UPs) also known as *Traffic Categories* (TCs). *Table 1* illustrates the various user priorities, their equivalent Access Categories and the typical type of data that these priorities are used.

Priority	User priority in 802.1D	Access category (AC)	Designation (informative)
Lowest	1	AC[0]	Background
	2	AC[0]	Background
	0	AC[1]	Best Effort
	3	AC[1]	Video
	4	AC[2]	Video
	5	AC[2]	Video
	6	AC[3]	Voice
Highest	7	AC[3]	Voice

Table 1. Priority classes in IEEE 802.11e

Each Access Category (AC) has its own parameters: *CWmin* (Minimum contention window), *CWmax* (Maximum Contention window), *AIFS* (Arbitrary Inter-frame Space which replaces DIFS of the conventional DCF), and *TXOPLimit* (Transmission Opportunity Limit)

The different Access Categories have different priority assigned by setting the above mentioned parameters. Contention is within a QSTA between ACs, as well as between QSTAs for medium. Thus, the eight different queues for the eight Traffic Categories within a station can be thought of as having eight queues inside a station along with QoS parameters that decide their priority. If two traffic categories have their counters zero at the same time, then we say that there has been a collision. Such a collision is also called a *virtual collision* because it is not an actual collision on the air channel, but it has happened virtually. Nevertheless, the back-off procedure is still executed. Figure 9 shows the traffic categories and the different parameters associated:
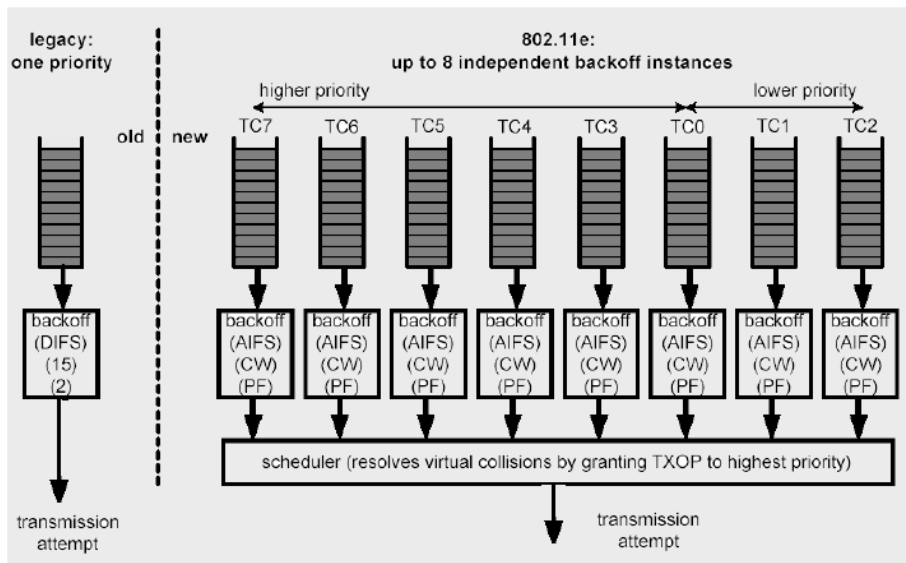
Figure 9. Queues for different Traffic Categories

Figure 10 shows the operation of different Traffic Categories within a QSTA. Each TC has its own AIFS and backoff window sizes. Thus, we have introduced priority between the different traffic types.
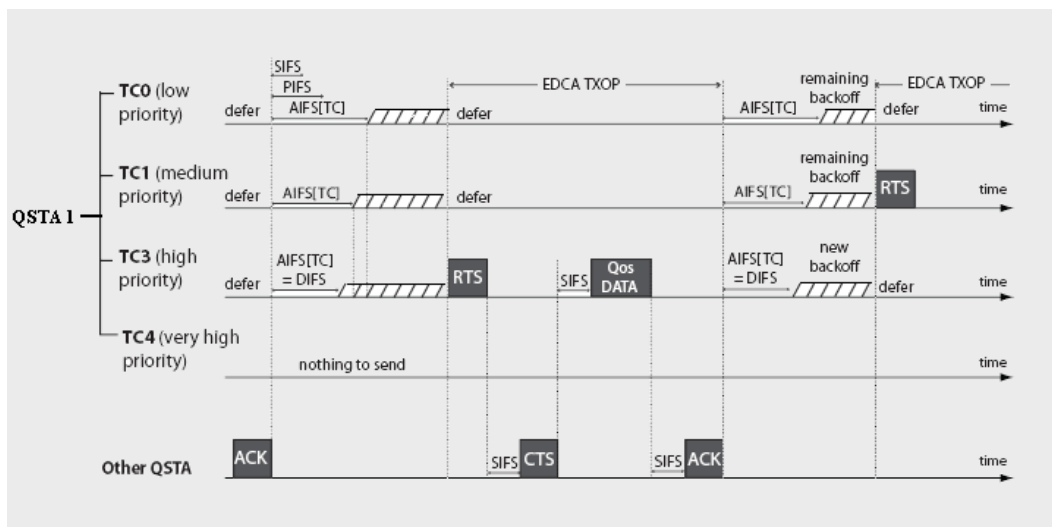

Figure 10. Traffic Categories within a station contending for access

It should be noticed that EDCA works well if the network is not heavily loaded. However, it has been shown that if the network is heavily loaded, then EDCA still provides expected priority for high priority flows, but at the cost of worst performance for lower priority flows. The HCCA mechanism though not very challenging, but is a promising approach and is discussed next.

b) HCF controlled Coordination Function (HCCA)

HCCA is the contention-free mode of the HCF. A central entity governs access to the channel. Stations register for access to channel by giving the Traffic Specs of the different streams that they contain and the QoS Access Point (QAP) polls stations to send their data. The beacon delay problem is

solved by not allowing a station to transmit a frame if the transmission will not finish before the next beacon. To address the problem of unlimited transmission time of polled stations, a variable called the TXOPlimit is introduced, which limits the transmission opportunity of the polled station. The QAP can start *Controlled Access Phases* (i.e. Contention-Free Periods) even during the contention period by using the HCCA prioritized access mechanism. The PIFS interval is shorter than AIFS or DIFS. A TCAPlimit is defined to limit the TXOPs during a beacon interval and to allow enough space for EDCF.

Besides these, IEEE 802.11e also introduces two optimizations techniques that can significantly improve performance: *Block acknowledgement* and *Direct link protocol*.

Block acknowledgement Mode can be negotiated between the QSTA and the QAP. Once the Block ACK mode has been established, multiple QoS frames are transmitted with SIFS separating them. This optimizes bandwidth as we don't need ACK for every frame. A single frame is used to acknowledge the reception of multiple frames.

Direct Link Protocol introduces the possibility of transmitting frames from one QSTA to another belonging to the same BSS directly, without the need for passing through the QAP. Normally, two QSTAs are not allowed to exchange frames directly with each other in a BSS, but only through the QAP. To enhance the performance for cases when the two QSTAs are in the same BSS and have some high bandwidth demanding applications running, then the DLP facilitates the efficient use of the transmission medium by allowing the two stations to communicate directly.

c) IEEE 802.11e Reference Scheduler and admission control

Up to eight different traffic classes of different priorities are allowed in the station. Before any traffic for a stream flows, the station sends a QoS request frame to the QAP carrying the *Traffic Specifications* (TSpecs) of the traffic stream. The *Minimum Data Rate*, *Mean Data Rate*, *Peak Data Rate*, *Burst Size*, *Minimum PHY Rate*, *Minimum Service Interval, Maximum Service Interval* and *Delay bound* fields in the TSpec express the characteristics of the traffic stream. When the QAP has received these requests, it assigns its Scheduler to calculate service schedule and the TXOP for each given SI (Service Interval). The calculation of the scheduled service interval is done as follows: first, the service scheduler calculates the minimum of "Maximum SI" of all the admitted streams. Let this minimum be *m*. Second, the scheduler calculates a number smaller than *m* and which is a sub-multiple of the beacon interval. This is the scheduled SI for all QSTAs with admitted streams. This same process is repeated as long as the maximum SI of the newly admitted stream is larger than the current scheduled service interval. If, however, the maximum SI of the newly admitted stream is smaller than the scheduled service interval, the scheduler needs to change the scheduled service interval to a value smaller than the maximum SI of this stream. The TXOPs also need to be recalculated with the new scheduled SI. If a stream is dropped, 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 option might require the announcement of a new schedule to all QSTAs. Different modifications can be implemented to improve the performance of the minimum scheduler. For example, a scheduler might generate different scheduled SIs for different QSTAs, and/or a scheduler might consider accommodating retransmissions while allocating TXOP durations.

When a new stream requests admission, the admission control process is done in three steps. First, the ACU (Admission Control Unit) calculates the number of frames that arrive at the mean data rate during the scheduled SI for this stream. Second, the ACU calculates the TXOP duration that needs to be allocated for the stream. Third, the ACU performs a final check before accepting the new flow to be sure that all the admitted streams have guaranteed access to the channel.

d) QoS for VBR traffic in IEEE 802.11 Networks

Providing QoS to applications is a challenging yet necessary task. In the case of wireless networks, the channel is prone to data loss, corrupted packets, high latency and jitter. For the wireless scenario, the MAC layer is the most important level for achieving this goal. The QoS can be characterized as prioritized or parameterized. Parameterized QoS is a strict requirement with quantitative bounds such as delay bound, data rate and jitter bound. One of the major weaknesses of

IEEE 802.11e is the inability of its scheduler to handle diverse applications, particularly VBR multimedia applications. It must be noted that this is a serious limitation because today, multimedia applications with VBR characteristics form a large portion of the traffic. In the following, we discuss the limitations of IEEE 802.11 and IEEE 802.11e.

– *QoS limitations of IEEE 802.11 DCF*: During the DCF mode, all stations contend to access the medium and there is no concept of priority. The DCF mode of the IEEE 802.11 fails to provide the desired bandwidth, delay or jitter requirements. The DCF provides only for best-effort services. Therefore for multimedia applications that depend heavily on delay bounds, the performance of DCF is unacceptable.

– *QoS limitations of IEEE 802.11 PCF*: The PCF mode attempts to provide QoS by using a centralized scheduler and polling frames, thus avoiding collisions that occurred due to distributed access in the DCF mode. It was meant to provide QoS to time-bounded multimedia applications. However, there are at least three main problems with the PCF approach that render it unable to provide the required QoS:

   Inefficient central polling scheme: All the communication between two stations in the same BSS has to be through the central Access Point. When the amount of traffic increases, this kind of mechanism proves quite wasteful for the bandwidth as the stations are better off exchanging directly.

   o *Unpredictable beacon delays*: The alternating CP and CFP may result in the Beacon Delay problem. At TBTT (Target Beacon Transmission Time), a PC schedules the beacon as the next frame to be transmitted. The beacon can be transmitted when the medium has been idle for at least PIFS. However, if at this time (TBTT), if the medium is busy due to some unfinished transmissions, the time bounded frames in CFP will suffer. The standard allows stations in Contention Period to transmit even if the transmission will not finish before TBTT. Thus the CFP period starts late and some or all flows in it suffer.

   o *Unknown transmission time of polled stations*: A station that has been polled by the PC can send a single frame up to max of 2304 bytes, and that frame may be fragmented. Further, different modulation and coding schemes are specified in IEEE 802.11a, thus after polling a station, the PC cannot tell how long will it take for this station to finish its transmission. This disables any attempt to provide QoS to the remaining stations that are to be polled during the CFP.

e) QoS limitations of IEEE 802.11e for Variable bit-rate Traffic

To provide QoS to applications in the wireless environment, the IEEE 802.11e standard has been proposed. However, the IEEE 802.11e is far from perfect. In certain scenarios and for certain types of applications, it fails to provide QoS guarantees. In the proceeding paragraph, we discuss the limitations of the IEEE 802.11e standard and in particular the centralized scheduler.

In the HCCA mode, stations send reservation requests to the scheduler and based on the information provided in the reservation request about the flow, the ACU admits or rejects the flow. Then again, based on the information about the characteristics of the flow, the scheduler allocates TXOPs to the stations and polls stations in a simple round robin manner. However, the information on which the scheduler bases time allocation to stations is the averaged information about the characteristics of the flows. The assumption is that these traffic flows are constant bit-rate and will continue this trend. The transmission duration of a station always stays the same in the reference scheduler and corresponds to the transmission time of an average sized packet or a burst of average size (whichever takes longer time) at the minimum physical rate. When the traffics are variable bit-rate, the performance of the scheduler degrades and it cannot guarantee QoS to the applications.

Several solutions have been proposed to deal with these limitations of IEEE 802.11e ([MAN02], [NI04], [NI05], [RAM05]). Among the proposed solution, we can mention:

- Dynamic allocation of additional time to VBR flows: The Access Point gathers queue information from stations to see if variations from the traffic reservation have caused queue build-up at QSTAs and if so, then it allocates additional time to prevent high delays.

- Variable service intervals: The authors of this solution use the concept of token bucket of time units or a TXOP timer for each QSTA. The transmission of a long burst in several SI spaced intervals leads to poor performance, but if the QSTA can use up its accumulated tokens at such needed instants, such bursts of data can be handled without degrading quality. The TXOP timer of a station increases at a constant rate of TD/mSI where TD is its transmission duration and the mSI is its minimum Service Interval. The time that the station used in its TXOP is then deducted from its TXOP timer. Thus over the long interval, a QSTA still has its TXOP equal to its mean data rate but it has more flexibility.

- Flexible HCF (FHCF): This approach handles the dynamic variation of the flows by adjusting the TXOPs of these flows, based on queue length estimations. It aims to be fair for both CBR and VBR traffic. There are two schedulers, the QAP scheduler and the node scheduler. The QAP scheduler calculates the ideal queue length of the TS (Traffic Stream) queues for each QSTA at the beginning of the next SI. When a QSTA sends a QoS data packet, the QAP uses the QoS control field of the IEEE 802.11e header to record its queue length at the end of TXOP. The QAP knows the time at which this TXOP ends, and hence using the queue length at the end of TXOP, the time at the end of TXOP and using other parameters like frame size and application data rate, the QAP scheduler estimates the queue length of a TS at the beginning of the next SI. To handle the case for VBR applications, the FHCF scheme uses a window of *w* already known real queue measurements to adjust the estimation. Then, the QAP scheduler compares the estimated queue length to the ideal queue length and reallocates the TXOP accordingly. The node scheduler also performs almost the same calculations. It has the task of redistributing the additional allocated time to the different TSs within the node.