A Survey of QoS Techniques and Enhancements for IEEE 802.11 Wireless LANs

Lassaâd Gannoune*, Stephan Robert* and Daniel Rodellar**

*Lassaad.Gannoune@eivd.ch, and Stephan.Robert@eivd.ch
Tel: +41 24 4239783 Tel: +41 79 5679835
Ecole d'Ingénieurs du Canton de Vaud (EIVD)
Rue de Cheseaux 1, 1400 Yverdon-les-Bains,
Switzerland.

**Daniel.Rodellar@swisscom.com
Tel: +41 31 8928587
Swisscom AG, Bern
Innovations
Broadband Networks
Ostermundigenstr. 93, 3050 Bern
Switzerland

Abbreviations and Acronyms

AC	Access Category	802.11e
ACK	Acknowledgement	
AIFS	Arbitration IFS	802.11e
CA	Collision Avoidance	
CAP	Controlled Access Period	802.11e
CC	Controlled Contention	802.11e
CCI	Controlled Contention Interval	802.11e
CCOP	Controlled Contention Opportunity	802.11e
CF-End	Contention-Free End	
CF-Poll	Contention-Free Poll	
CFB	Contention Free Burst	802.11e
CFP	Contention Free Period	
CP	Contention Period	
CSMA	Carrier Sense Multiple Access	
CW	Contention Window	
CWmax	Contention Window Maximum	
CWmin	Contention Window Minimum	
DCF	Distributed Coordination Function	
DIFS	DCF Inter-Frame Space	
EDCF	Enhanced DCF	802.11e
EIFS	Extended Inter-Frame Space	
HC	Hybrid Coordinator	802.11e
HCF	Hybrid Coordination Function	802.11e
ISM	Industrial, Scientific and Medical	
MAC	Medium access	
MPDU	MAC Protocol Data Unit	
MSDU	MAC Service Data Unit	
NAV	Network Allocation Vector	
PC	Point Coordinator	
PCF	Point Coordination Function	
PF	Persistence Factor	802.11e
PHY mode	Physical layer mode	
PIFS	PCF Inter Frame Space	
QAP	QoS-supporting Access Point	802.11e
QBSS	QoS basic Service Set	802.11e
QSTA	QoS Station	802.11e
RR	Reservation Request	802.11e
RTS/CTS	Request to Send/Clear to Send	
SIFS	Shortest Inter Frame Space	
STA	Station	
TBTT	Target Beacon Transmission Time	
TC	Traffic Category	802.11e
TID	Traffic Identifier	802.11e
TS	Traffic Stream	802.11e
TSPEC	Traffic Specification	802.11e
TXOP	Transmission Opportunity	802.11e
WLAN	Wireless Local Area Network	
WM	Wireless Medium	
WSTA	Wireless QoS Station	802.11e

List of Tables

List of figures

Table of Contents

1 Introduction

Wireless communication technology has gained widespread acceptance in recent years. Wireless Local Area Networks (WLANs) are more and more commonly and widely used, since the advent of the IEEE 802.11 WLAN standard [1]. The main characteristics of the 802.11 WLAN technology are simplicity, flexibility and cost effectiveness. This technology provides people with ubiquitous communication and computing environment in offices, hospitals, campuses, factories, airports and stock markets. Simultaneously, people start looking to their laptops or PDAs to deliver the broadband multimedia applications currently being developed. These applications include streaming media, interactive collaboration, videoconferencing and downloadable content such as multimedia messaging. However, multimedia applications require a certain quality of service (QoS) support such as guaranteed bandwidth, delay, jitter and error rate. Guaranteeing those QoS requirements is a challenging task with regard to 802.11 WLAN protocols and Medium Access Control (MAC) functions.

While QoS issues in the Ethernet are not been considered as a high priority topic due to the huge improvements of physical layer bandwidths, the IEEE 802.11e group [21] is developing MAC enhancements to support QoS sensitive applications (e.g., multimedia applications). These enhancements will enable a better mobile user experience and will make more efficient use of the wireless channel.

The purpose of this report, in a first part, is to provide a comprehensive view of the 802.11 WLAN and analyze its QoS limitations and issues. This report, also surveys the current research concerned with the problem of providing QoS in 802.11 wireless networks. The main QoS techniques and enhancements for 802.11 presented in this report will be an integral part of the upcoming 802.11e standard.

In a first section (section 2), we give an overview of the main wireless LAN MAC protocols. We describe the main MAC protocols and techniques as well as their different features. Then, some throughput issues related to WLAN are discussed. Traditional QoS features and limitations for 802.11 standards are addressed in section 3. The section 4 surveys the research efforts that intend to provide new mechanisms for QoS support in 802.11 WLAN. Then, we detail the main new designed access methods in the context of the upcoming 802.11e standard. We also present the performance of these new methods. In section 5 we address the main issues still unresolved in 802.11e and we propose some research items as basis for future work. Finally, in section 6 we give some concluding remarks.

2 Overview of WLAN protocols and topologies

In order to have a comprehensive view about wireless LAN protocols for 802.11, a first section surveys the main channel access mechanisms designed for radio channels. Then, different Medium Access Control (MAC) techniques are presented as well as the different related wireless topologies. Finally, this section describes and discusses the throughput considerations and issues at the MAC layer.

2.1 Channel access mechanisms (MAC protocols)

The main role of the MAC protocol is to regulate the usage of the medium and this is done through a channel access mechanism. A channel access mechanism is a way to divide the main resource, the radio channel, between nodes, by regulating the use of it. This access mechanism tells each node when it can transmit and when it is expected to receive data. The channel access mechanism is the core of the MAC protocol. In this section, we describe *TDMA* (Time Division Multiple Access), *CSMA* (Carrier Sense Multiple Access) and *polling* techniques which are the three main classes of channel access mechanisms for radio.

2.1.1 TDMA

In this section, TDMA (Time Division Multiple Access) is described as a channel access mechanism. TDMA is a simple protocol. A specific node, the **base station**, has the responsibility to coordinate the nodes of the network. The time on the channel is divided on *time slots*, which are generally of fixed size. Each node of the network is allocated a certain number of slots when it can transmit. Slots are usually organised on a frame (*superframe*), which is repeated on a regular basis. Figure 1 illustrates the TDMA channel access mechanism.

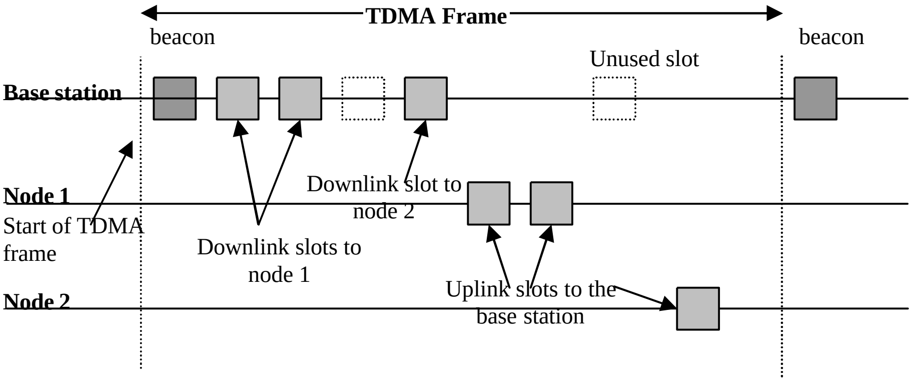

Figure 1. TDMA channel access mechanism

The superframe begins by a specific frame, the beacon (a management frame), in which the base station specifies the organisation of the frame. Each node just needs to follows blindly the instruction of the base station. The frame is organised as downlink (i.e., base station to node) and uplink (i.e., node to base station) slots, and all the communications goes through the base station. A service slot allows a node to request the allocation of a connection, by sending a connection request message in it. In some standards, uplink and downlink frames are on different frequencies, and the service slots might also be a separate channel.

TDMA suits very well phone applications, since those applications have very predictable needs (fixed and identical bit rate, CBR traffic). Hence, each handset is allocated a downlink and an uplink slot of a fixed size (the size of the voice date for the duration of the frame). Cellular phone standards (GSM in Europe, TDMA and PCS in the USA) and cordless phone standards (DECT in Europe) use TDMA as their main channel access mechanism. In addition TDMA achieves low latency and guarantee of bandwidth (where CSMA/CA is quite bad).

In contrast to cellular phone applications, TDMA is not well suited for data networking applications, because it is very strict and inflexible. IP is connectionless and generates bursty traffic which is very unpredictable by nature, while TDMA is connection oriented (so it has to suffer the overhead of creating connections for single IP packets). Moreover, TDMA uses fixed size packets and usually symmetrical link, which does not suite IP (i.e., variable size packets).

TDMA is also very dependant of the quality of the frequency band. In a dedicated “clean” band, as it is the case for cellular phone standard, TDMA is very adequate. But, its inflexibility and its unawareness of what its happening on the channel, make TDMA unable of adapting to the bursty interference sources found in the unlicensed band.

2.1.2 CSMA/CA

CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance) is the channel access mechanism used by most wireless LANs in the Industrial Scientific and Medical (ISM) band. The basic principles of CSMA are to listen before talk and the contention. This is an asynchronous message passing mechanism (connectionless), delivering a best effort service, and no bandwidth and latency guarantee. CSMA is fundamentally different from the channel access mechanism used by cellular phone systems (i.e., TDMA).

CSMA/CA is derived from the channel access mechanism CSMA/CD (*Collision Detection*) employed by Ethernet. However, collisions waste valuable transmission capacity, so rather than the collision detection (CD) used in Ethernet, CSMA/CA uses collision avoidance. *Collision avoidance (CA)*; on a wire, the transceiver has the ability to listen while transmitting and so to detect collisions (with a wire all transmissions have approximately the same strength). But, even if a radio node could listen on the channel while transmitting, the strength of its own transmissions would mask all other signals on the air. Thus, the protocol can not directly detect collisions like with Ethernet and only tries to avoid them.

The 802.11 standard defines the Distributed Coordination Function (DCF) as its fundamental access method and is based on CSMA/CA. DCF allows each multiple independent nodes (or stations) to interact without central control. Figure 2 illustrates the basic access method used in the DCF protocol.

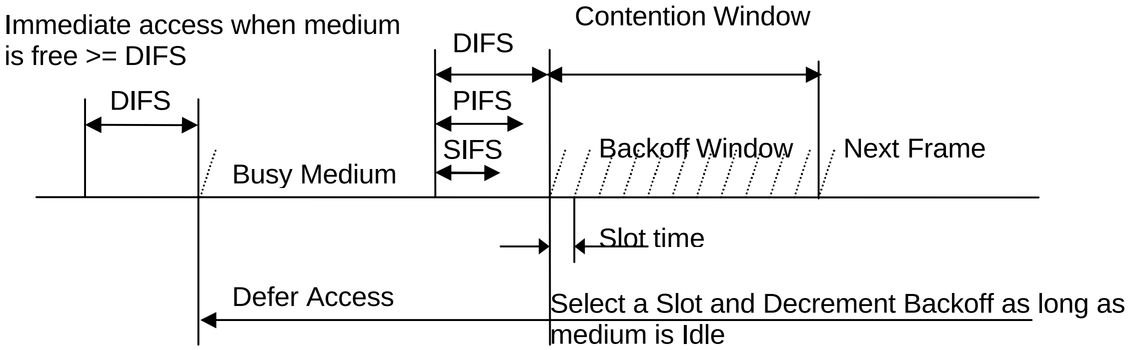

Figure 2. DCF (CSMA/CA) Basic access method

The DCF protocol on each station starts by listening on the channel (this is called carrier sense), and if it finds the channel idle for at least a DIFS (DCF Inter Frame Space) period, it sends the first packet in the transmit queue. If it is busy (either another node transmission or interference), the station waits the end of the current transmission and then starts the contention. It selects a random slot time, a so called backoff time, from a Contention Window (CW) and waits for DIFS and its backoff time. The backoff time is given by the following formula.

$$T_{backoff} = Rand(0, CW) \times T_{slot}$$

Where T_{slot} is a time slot specific to the physical layer [1] and Rand() is a uniform distribution random function. This back-off time is computed to initialize the back-off timer and this timer is only decreased when the medium is idle. When the medium is sensed to be busy, this timer is frozen.

When its back-off timer expires, and if the channel is still idle, the node sends the frame. Thus, the node having chosen the shortest backoff time wins and transmits its frame. The other nodes just wait for the next contention (after waiting for the end of this packet transmission). Because the contention period is derived from a random number chosen with a uniform distribution, and done for every frame, each station is given an equal chance to access the channel (statistically mean average).

As a feature for the radio link, collisions can not be detected and because the radio needs time to switch from receive to transmit, this contention is usually slotted. A transmission only starts at the beginning of a slot; that is 50µs in 802.11 Frequency Hopping (FH) and 20 µs in 802.11 Direct Sequence (DS). This makes the average contention delay larger, but reduces significantly the collisions.

As with traditional Ethernet, the inter-frame spacing plays a major role in coordinating access to the transmission medium. 802.11 CSMA/CA protocol (in DCF mode) uses four different inter-frame spaces. The relationship between them is shown in figure 2 (i.e., SIFS, PIFS, DIFS and EIFS):

- Shortest Inter Frame Space (SIFS): the SIFS is used for the highest priority transmissions, such as RTS/CTS frames and positive acknowledgements. High priority transmissions can begin once the SIFS has elapsed.
- PCF Inter Frame Space (PIFS): The PIFS is used by the PCF (Point Coordination Function) during contention-free operation (In addition to DCF, PCF is another access mode provided by 802.11). Stations with data to send in the contention-free period can transmit after the PIFS has elapsed and thus pre-empt any contention-based traffic.
- DCF Inter Frame Space (DIFS): The DIFS is the minimum idle time for contention-based access. Stations may have immediate access to the medium if it has been free for longer than DIFS.
- Extended Inter Frame Space (EIFS): The EIFS is not illustrated in the figure 2, because it is used only when there is an error in the frame transmission.

Varying inter-frame spacing create different priority levels for different types of traffic. The logic behind this is simple; high-priority traffic does not have to wait as long after the medium has become idle. These different inter-frame spaces depend on physical layer parameters and hence different physical layers can specify different inter-frame space times.

The main advantages of CSMA/CA are that it is suited for network protocols such as TCP/IP, adapts quite well with the variable condition of traffic and is robust against interferences.

2.1.3 Polling

Polling is the third major channel access mechanism, after TDMA and CSMA/CA for radio channels. The 802.11 wireless standard offers a polling channel access mechanism (Point Coordination Function) in addition to CSMA/CA.

Polling is in fact in between TDMA and CSMA/CA. The base station retains total control over the channel, but the frame content is no more fixed, allowing variable size packets to be sent. The base station sends a specific packet, a poll packet, to trigger the transmission by the station. The station just waits to receive a poll packet, and upon reception sends what it has to transmit. Figure 3 shows the polling access scheme with a base station and two mobile stations.

Polling can be implemented as a connection oriented service (like TDMA, but with higher flexibility in packet size) or connectionless service (asynchronous packet transmission). The base station can either poll permanently all the stations of the network to check if they have something to send (this is adequate with a very limited number of stations), or the protocol use reservation slots where each station can request a connection or to transmit a packet (depending on the nature of MAC protocol connection oriented or not).

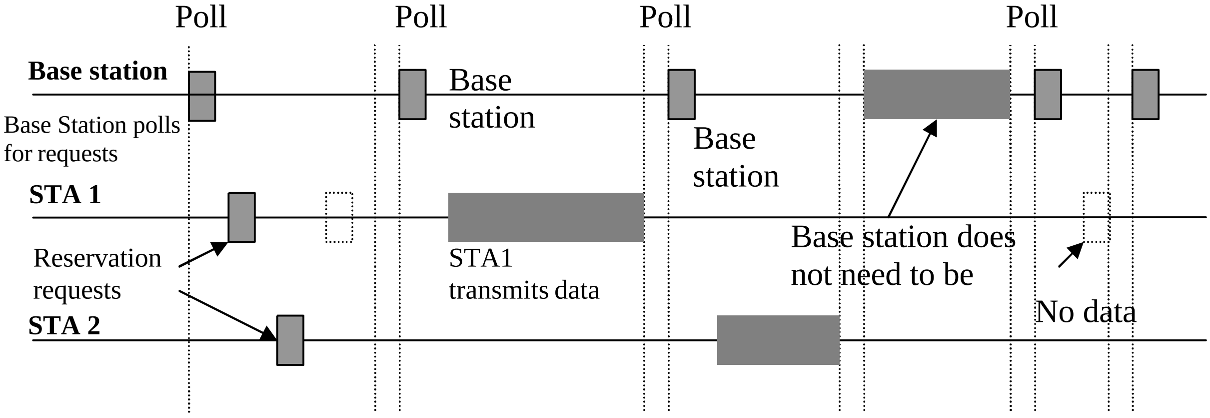

Figure 3. Polling channel access mechanism

The most successful networking standard using polling is IEEE 802.12 in which the polling mechanism does not use any bandwidth (its done out of bandwidth through tones), leading to a very efficient use of the channel. In the case of 802.11 wireless networks, all the polling packets are sent over the air link, generating much more overhead. Furthermore, polling in the 802.11b standard does not use any reservation mechanism, and thus it is inflexible and generates a significant overhead [19].

As CSMA/CA offers ad-hoc networking (no need of a base station) and similar performance, it is usually preferred in most wireless LANs. Consequently, more 802.11b vendors prefer to implement the DCF mode (CSMA/CA) over the coordinated mode.

In the section 2.5, we describe the PCF access method which is main based on a polling mechanism designed in the context of 802.11 networks which is aims to provide quality of service (QoS).

2.2 MAC techniques

In section 2.1 we have described the major channel access mechanisms and especially the main principle of CSMA/CA, but most MAC protocols use additional techniques to improve the performance of CSMA/CA. In this section, we focus on the main techniques used in the context of 802.11 wireless networks with CSMA/CA.

2.2.1 MAC retransmissions (vs Error recovery with DCF)

As described in the previous section, the main problem of CSMA/CA is that the transmitter can not detect collisions on the medium. There is also a higher error rate on the air than on a wire, so a higher chance of frames being corrupted. TCP does not like packet losses at the MAC layer. Because of that, most MAC protocols implement *positive acknowledgement* and *MAC level retransmissions* to avoid loosing packets on the air.

The principle is quite simple; each time a station receives a frame, it sends back immediately a short message, an acknowledgement (i.e., ACK) to the transmitter indicating that it has successfully received the frame without errors. In some cases, the sender can infer a frame loss by the lack of a positive acknowledgement from the receiver. So it will retransmit the packet (after contending again for the medium). The contention window size (starting from a predefined CW_{min}) increases when a transmission fails, thus another backoff time is computed with a doubled size of the CW.

The purpose of the increase of the contention window is to reduce the number of collisions, after each unsuccessful transmission. The CW is doubled until a maximum predefined value CW_{max}.

Retry counters are incremented when frames are retransmitted. Each frame or fragment has a single retry counter associated with it. Stations have two retry counters, the *short retry count* and the *long retry count*. Depending on the length of the frame, it is associated with either a short or a long retry counter. Indeed, frames that are shorter than the RTS threshold (the next section will give more details on fragmentation conditions) are considered to be short while frames longer than the threshold are long.

Frames retry counts begin at 0 and are incremented when a frame transmission fails. The short retry count is reset to 0 when :

- A CTS frame is received in response to a transmitted RTS
- A MAC –layer acknowledgement is received after a non fragmented transmission
- A broadcast or multicast frame is received

The long retry count is reset to 0 when :

- A MAC-layer acknowledgement is received for a frame longer than the RTS threshold
- A broadcast or multicast frame is received

In addition to the associated retry count, fragments are given a maximum *lifetime* by the MAC. When the first fragment is transmitted, the lifetime counter is started. When the lifetime limit is reached, the frame is discarded and no attempt is made to transmit any remaining fragments.

When a station transmits a frame, it must receive an acknowledgement from the receiver or it will consider the transmission to have failed. Failed transmissions increment the retry counter associated with the frame (or fragment). If the retry limit is reached, the frame is discarded and its loss is reported to higher layer protocols.

2.2.2 Fragmentation

The radio medium has a higher *error rate* than the wire. In the previous section MAC level retransmissions are described in order to cope with frame losses and thus to provide an error recovery mechanism.

MAC level retransmissions solve the problem of frame losses, but it is not performant. Indeed, if the packet to retransmit is long and contains only one error, the station needs to retransmit it entirely. When the error rate is significantly high, it could be possible that the probability of error in large frames is dangerously close to 1 (the station can not fit the frame between the bursts of errors due to fading or interferers) and this packet can not be delivered.

In order to improve reliability in the presence of interference, 802.11 networks use *fragmentation* which break large frames into small pieces to fit through the wireless channel. Fragmentation takes place when a higher-level packets length exceeds the fragmentation threshold (configured by the network administrator). All fragments have the same sequence number, but they have ascending fragment numbers to help in reassembly. Frame control information indicates whether more fragments are coming. All the fragments that comprise a frame are normally sent in a fragmentation *burst*, which is shown in figure 4. This figure also incorporates an RTS (request to send)/CTS (clear to send) exchange which enables to prevent collisions. The next section (section 2.2.3) presents in more detail the RTS/CTS exchange procedure. The figure also shows how the NAV (Network Allocation Vector) and SIFS are used to control the access to the medium.

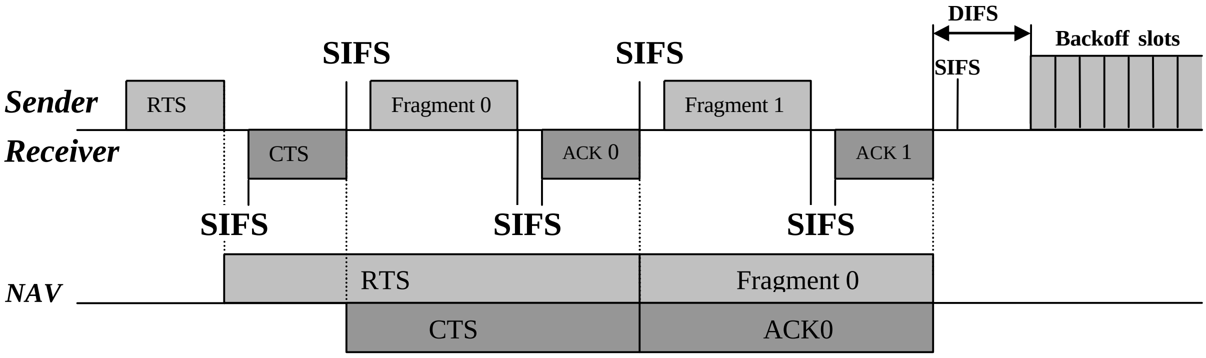

Figure 4. A frame sent in a fragmentation burst

The NAV is used to implement a *virtual carrier sensing function*. In fact, the NAV is a timer that indicates the amount of time the medium will be reserved. A station (sender) sets the NAV to the time for which it expect to use the medium. All other stations use count down from the NAV to zero. When the NAV is nonzero, the virtual carrier sensing function indicates that the medium is busy and when it reaches zero, the virtual carrier sensing function indicates that the medium is idle.

As shown in figure 4, the fragments and their acknowledgements are separated by SIFS, so a station retains control of the channel during a fragmentation burst. The NAV is then used to ensure that other stations don’t use the channel during the fragmentation burst.

Each fragment is individually checked and retransmitted if necessary. The first advantage is that in case of error, the mobile station needs only to retransmit one small fragment, so it is faster. The second advantage is that if the medium is very noisy, a small frame has a higher probability to get through without errors, so the mobile station increases its chance of success in bad conditions.

2.2.3 RTS/CTS (hidden node problem)

The main effect of transmission on radio waves is the attenuation of the signal. Because of this attenuation a problem of *hidden nodes* appears. The hidden node problem comes from the fact that all nodes may not hear each other because the attenuation is too strong between them. Since transmissions are based on the carrier sense, those nodes ignore each other and may transmit at the same time. This is usually a good situation and allows frequency re-use (these nodes are effectively in different cells). However, for a node placed in between, these simultaneous transmissions have a comparable strength and so collide (in its receiver). Due to those collisions, this node could be unreachable.

Furthermore, collisions resulting from hidden nodes may be hard to detect because wireless transceivers are generally half-duplex; they do not transmit and receive at the same time. In order to prevent collisions, 802.11 wireless networks allows stations to use Request to Send (RTS) and Clear to Send (CTS) signals to clear out an area. Figure illustrates the hidden node problem and the RTS/CTS procedure.

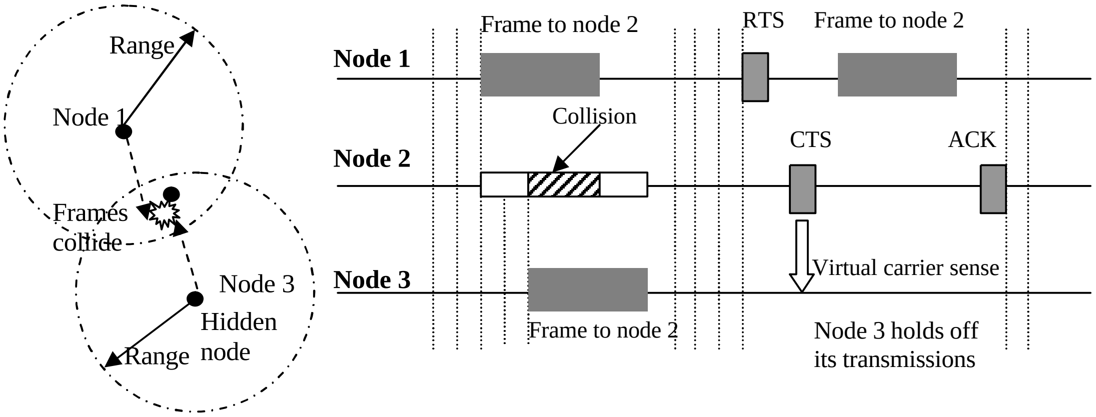

Figure 5. Hidden nodes and RTS/CTS procedure: node 3 is hidden to node 1

RTS/CTS is a handshaking; before sending a packet, the transmitter sends a RTS frame and waits for a CTS frame from the receiver. The reception of CTS indicates that the receiver is able to receive the RTS, so the frame (i.e., the channel) is clear in its area.

The RTS frame serves several purposes; in addition to reserving the radio link for transmission, it silences any stations that hear it. Like the RTS frame, the CTA frame silences stations in the immediate vicinity. Once the RTS/CTS exchange is complete node 1 can transmit its frames without worry of interference from any hidden nodes. Indeed, hidden nodes beyond the range of the sending station (i.e., node 3) are silenced by the CTS from the receiver (by setting their NAV to the duration of the frame exchange). When the RTS/CTS clearing procedure is used, any frames must be positively acknowledged.

The multi-frame RTS/CTS transmission procedure consumes a fair amount of capacity, because of additional latency incurred before transmission can start. As a result, its only used in high capacity environments and environments with significant contention on transmission. Thus, it is not used for small frames or lightly loaded environments.

It is possible to control the RTS/CTS procedure by setting the RTS *threshold* if the device driver for the 802.11 network card allows to adjust it. Hence, the RTS/CTS exchange is performed for frames larger than the threshold and frames shorter than the threshold are simply sent.

2.3 Network Topology of WLANs

The topology of wireless LAN is different from traditional LANs. The connectivity is limited by the range, so we usually do not have complete coverage (some nodes may not see each other). This breaks some assumptions of higher layers. In order to overcome this, either the network is divided in cells managed by an *Access Point*, or the network use *MAC level forwarding (such as in HyperLan).*

The basic building block of an 802.11 network is the basic service set (BSS), which is simply a group of stations that communicate with each other. Communications take place within an area called the basic service area, defined by the propagation characteristics of the wireless medium. When a station is in the basic service area, it can communicate with the other members of the BSS. Two types of BSSs are defined.

Independent networks (Ad-hoc networks)

Stations in an independent BSS (IBSS) communicate directly with each other and thus they must be within direct communication range and they ust have no infrastructure (like no Network Provider). The smallest possible 802.11 network is an IBSS with two stations. IBSSs are typically composed of small number of stations set up for a specific purpose and for a short period of time. One common use is to create a short-lived network to support a single meeting in a conference room. IBSSs are referred to as ad-hoc BSSs or ad-hoc networks.

Infrastructure networks

Infrastructure networks are distinguished by the use of an *Access Point.* Access points are used for all communications in infrastructure networks, including communication between mobile nodes in the same service area. If one mobile station in infrastructure BSS needs to communicate with a second mobile station, the communication must take two hops. First, the station sender transfers the frame to the access point. Second, the access point transfers the frame to the station destination. Although the multihop transmission takes more transmission capacity than a direct frame from the sender to the receiver, it has two major advantages:

- An infrastructure BSS is defined by the distance from the access point. All mobile stations are required to be within reach of the access point, but no restriction is placed on distance between stations themselves. Direct communications between stations would save transmission capacity but at the cost of increased physical layer complexity because stations would need to maintain neighbour relationships with all other stations within the service area.

- Access points in infrastructure networks allow stations to save power. Indeed, access points can note when a station enters into a power saving mode and buffer frames for it. Battery-operated stations can turn the wireless transceiver off and power it up only to transmit and retrieve buffered frames from the access point.

In an infrastructure network, stations must be associated with an access point to obtain network services. Association is the process that enables a station to join an 802.11 network. Stations always initiate the association process, and access points may choose to grant or deny access based on the contents of an associated request. A station can be associated with only one access point. However, 801.11 standard places no limit on the number of stations that an access point may serve. In practice, however, the relatively low throughput of wireless networks is far more likely to limit the number of stations placed on a wireless network.

Extended service areas

BSSs can create coverage in small offices and homes, but they can not provide by themselves network coverage to larger areas. 802.11 allows wireless networks of arbitrarily large sizes to be created by linking BSSs into an extended service set (ESS). An ESS is created by chaining BSSs together with a backbone network (such as Ethernet).

Stations within the same ESS may communicate with each other, even though these stations may be in different basic service areas. For stations in an ESS to communicate with each other, the wireless medium must act like a single layer 2 connection. Access points act as bridges, and the backbone network plays the role of a layer 2 connection. A router on the backbone network can be used to allow connections from stations to the Internet. In this case, the router must use a single MAC address to deliver frames to a station; the access point with which that station is associated delivers the frame. The router remains ignorant of the location of the station and relies on the access point to deliver the frame.

2.4 Throughput issues

Bit rate versus maximum user throughput

Like for wired products, most radio LAN vendors indicate only the bit rate of their products, called signalling rate. For example, Ethernet is 10 Mbps, 100Mbps, or 1 Gbps, and most radio LAN products, 1Mbps, 2Mpbs, 5.5 Mbps and higher rate at up to 11 Mbps (802.11b products). The signalling rate is the speed at which bits are transmitted over the medium, but because of the many overheads of the protocols used, the user throughput is usually less. Wireless LANs protocols have a *higher overhead* than their wired counterpart (such as Ethernet). This higher overhead is due to different factors:

- The first is the radio technology: radio receivers require large synchronization fields (receiver training, antenna selection, etc...). In addition the radio itself is slow to react (switch from receive to transmit), so needs large slots in the contention window and between the transmissions of packets.

- The second is the addition of the features necessary to the radio protocol which makes the frame MAC headers larger (fields for network identification, encryption parameters, etc...) or introduces new management frames (e.g., synchronization, authentication, association).

- The third is related to the improvement of reliability. For example, big packets may be split into small independent fragments to decrease the error probability. Acknowledgements and RTS/CTS frames add also some overhead. Having a slotted contention decreases the amount of collisions but increases the average contention delay.

Adding all of these factors, makes a significant difference between the signalling rate and the user throughput. In the case of Ethernet it is possible to reach 80% to 90% of the signalling rate. While, for the most radio products the user throughput is usually between 50% and 70% of the signalling rate [19].

Multi-rate system considerations

Multi-rate systems also impact the overhead of the system especially at high rates. All the basic parts of the MAC protocol (headers, management frames, contention window) is designed for slowest rate, so when going to higher rates their relative size increase (their duration remain the same while the payload duration decreases).

For example, when we double the signalling rate to send 1500 bytes with 802.11, the overhead of the contention window double, the overhead the MAC level acknowledgements and RTS/CTS double and the overhead of the header increases by 28%. Lucent claims that increasing the bit rate from 2 to 10 Mbps (Lucent turbo PPM DS modulation), the effective user throughput is increased only by a factor of 3[19].

Contention and congestion

When there is many nodes sending packets on the network, the probability of having two nodes choosing the same slot in the contention window increases. In fact, when two nodes choose the same slot, their frames collide and are lost. This means when the level of *contention* increases, the number of retry increases, so the performance of the network drops to the point of *congestion*.

A solution to this problem is to use RTS/CTS procedure which makes each collision shorter (only on RTS/CTS frames). With RTS/CTS, 802.11 can support more than a dozen active nodes without significant reduction in performance due to contention (those nodes have to share the available bandwidth). The RTS/CTS handshake is done at basic rates, thus its benefit tends to decrease for the highest transmissions rates.

3 Traditional QoS mechanisms and limitations in 802.11 WLAN

In order to support applications that require near real-time services [20], 802.11 standard [1] includes a second coordination function (in addition to DCF) to provide a different way for accessing the wireless medium. The point coordination function (PCF) allows an 802.11 network to provide a fair access to the medium. The PCF uses both a polling access scheme and a DCF access based on CSMA/CA which alternates on time. Since PCF is an optional part of 802.11 specification [1], products are not required to implement it.

Contention-free service using PCF

The contention-free service is not provided full-time. Periods of contention free service arbitrated by the Point Coordinator (PC) alternate with the standard DCF-based access (or contention period). The duration of the contention free period can be configured. 802.11 describes the contention-free period as providing “near isochronous” service because the contention-free period will not always starts at the expected time (refer to the limited QoS support with PCF section below).

A contention-free service uses a centralized access control method. Access to the medium is restricted by the Point Coordinator, a specialized function implemented in access points. Associated stations can transmit data only when they are allowed to do so by the point coordinator. Contention-free access under the PCF resembles token-based networking protocols, with the point coordinator’s polling taking the place of a token. Despite, access is under the control of a central entity, all transmissions must be acknowledged. The figure 6 illustrates the PCF access method.

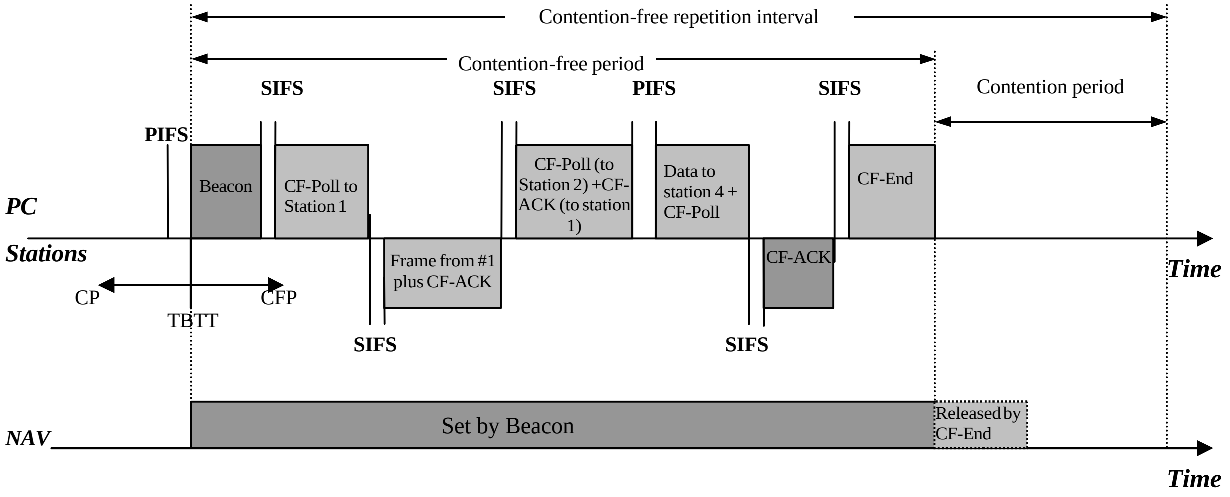

Figure 6. PCF access scheme

When the PCF is used, time on the medium is divided into contention-free period (CFP) and the contention period (CP). Access to the medium during the CFP is controlled by the PCF, while access to the medium in CP is controlled by the DCF and the rules defined in the section 2.1.2. In order to be “fair” with contending traffic, the contention period must be long enough for the transfer of at least one maximum size frame and its associated acknowledgement. Alternating periods of contention-free service and contention based service repeat at regular intervals, called the *contention-free repetition interval* (known also as a *superframe*). The figure 7 illustrates CFP and CP alternation.

At the beginning of the CFP, the PC (which resides in AP) transmits a management frame, called beacon. One of the beacon role component is the maximum duration, *CFPMaxDuration*, of the CFP. The PC generates beacons at regular beacon frame intervals, thus every station knows when the next beacon frame will arrive. This time is called target beacon transmission time (*TBTT*). All stations receiving the beacon set the NAV to the maximum duration to lock out DCF-based access to the wireless medium. The access point maintains a polling list of associated stations and polls any station in this list. Since time in the CFP is precious, acknowledgements, polling, data transfer may be combined to improve efficiency (as shown in figure 6).

All CFP transmissions are separated by short inter frame spaces (SIFS) and PCF inter frame spaces (PIFS), where PCF waits some time if ther is no response from a polled station, to prevent interference from DCF traffic (both are shorter than DCF interframe space, DIFS).

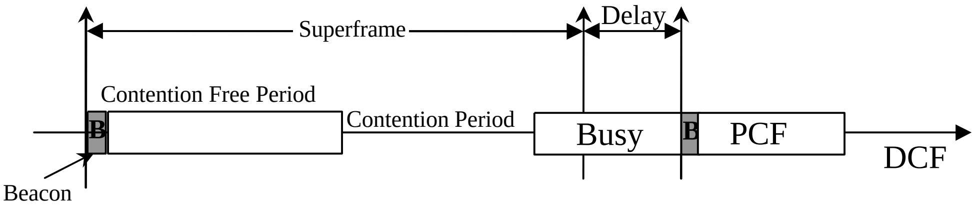

Figure 7. PCF and DCF alternation

Although the contention-free service is designed in 802.11 networks to provide QoS for real-time traffic, this service presents some limitations and can not meet the requirements of real-time traffic. In the following we describe the main limitations related to PCF.

Limited QoS support with PCF

The point co-ordination function (PCF) method has the following limitations:

- Unpredictable beacon delay : this problem is related to the uncontrolled length of the CP. Indeed, the minimum length of the CP is the time required to transmit and acknowledge one maximum size frame. It is possible for the contention service to overrun the end of the CP, due to the transmission of a contending traffic. When the contention based service runs past the TBTT, the CFP is *foreshortened,* and hence the beacon is delayed (the time the beacon frame is delayed, i.e., the duration it is sent after the TBTT). Figure 7 illustrates the delay of the beacon transmission. This also delays the transmission of time-bounded MSDUs that have to delivered in CFP. From the legacy 802.11 standard stations can continue their transmissions even if the MSDU delivery can not finish before the upcoming TBTT[3]. This may severely affect the QoS as this introduces unpredictable delay for each contention-free period.

- Unknown transmission time of polled stations : a station that has been polled by the PC is allowed to send a single frame that may be fragmented and of arbitrary length, up to the maximum of 2304 bytes (2312 bytes with encryption). Furthermore, different modulation and coding schemes are specified in 802.11a, thus the duration of the MSDU delivery that happens after the pooling is not under control of the PC. This may destroy any attempt to provide QoS to other stations that are polled during the rest of the CFP.

- No knowledge of the offered traffic at the stations : with CFP, the base station (AP) has no knowledge of the offered traffic at the polled stations. Thus, when polling the different stations with a round-robin scheduling algorithm, the PC may waist a lot of time until polling a special station having a critical time traffic (e.g., CBR traffic). This may affect the QoS parameters for these traffic categories. Hence, with PCF there is no efficient *scheduling algorithm*, which has the knowledge of the different traffic categories at associated stations and uses this knowledge in order to meet the requirements (e.g., *latency*, *bandwidth*) of these different traffic categories.

Finally, PCF with the round robin scheduling algorithm, performs better if every node has something to transmit than only one node does (in the latter case between each packet of the node, the protocol has to poll all other nodes of the network for nothing). But, PCF can not provide some level of assurance for consistent real-time traffic delivery. Due to these limitations, new mechanisms are designed and under standardisation in order to support QoS (on the air) for 802.11 wireless networks. The following section will detail the main concepts and mechanisms for MAC enhancements for supporting quality of service.

4 New QoS mechanisms and enhancements schemes for 802.11 wireless networks

There is more than one way to characterize Quality of Service (QoS). Generally, QoS is the ability for a network element (e.g., an application, a host or a router) to provide some level of assurance for consistent network data delivery [11]. Some applications are more stringent about their QoS requirements than others, and for this reason (among others), we have two basic approaches of QoS available:

- *Resource reservation* (integrated services approch) : network resources are apportioned according to an application’s QoS request, and subject to bandwidth management policy.

- *Prioritization* (differentiation services approach) : network traffic is classified and apportioned network resources according to bandwidth management policy criteria. To enable QoS, network elements give preferential treatment to classifications identified as having more demanding requirements.

These approaches of QoS can be applied to individual application “flows” or to flow aggregates, thus there are two other ways to characterize types of QoS:

- *Per flow* : a flow is defined as an individual, uni-directional, data stream between two applications (a sender and a receiver), uniquely identified by a 5-tuple: transport protocol, source address, source port number, destination address, and destination port number [11].

- *Per aggregate* : an aggregate is simply one or two flows. Typically, the flows will have something in common, e.g., any one or more of the 5-tuple parameters, a label or a priority number, or some authentication information.

In the following sections we firstly present the designed QoS techniques for the MAC DCF access method which are based on service differentiation approach. Secondly, the upcoming 802.11e standard is presented. The main 802.11e QoS enhancements are based both on a differentiated services and integrated services approaches.

4.1 Differentiated services based on DCF

In order to enhance the IEEE 802.11 DCF access method, the following techniques were designed in order to introduce priorities in DCF operation mode.

4.1.1 Backoff increase function (or Deng scheme)

Deng and Chang in [2] propose a method for service differentiation with minimal modification of the IEEE 802.11 standard [1]. This method uses two properties of 802.11 DCF to provide differentiation, the interframe space (IFS) used between data frames, as the waiting time before the contention window (or the backoff time) and the backoff mechanism. If two stations use different IFS, a station with the shorter IFS will get higher priority than a station with a longer IFS. Table 1 shows the four defined classes.

Priority	IFS	Backoff algorithm
0	DIFS	$B = \frac{2^{2+i}}{2} + \left\lfloor r \times \frac{2^{2+i}}{2} \right\rfloor$
1	DIFS	$B = \left\lfloor r \times \frac{2^{2+i}}{2} \right\rfloor$
2	PIFS	$B = \frac{2^{2+i}}{2} + \left\lfloor r \times \frac{2^{2+i}}{2} \right\rfloor$
3	PIFS	$B = \left\lfloor r \times \frac{2^{2+i}}{2} \right\rfloor$

Table 1. DENG priority classes

To further extend the number of available classes, different backoff algorithms are used depending on the priority class. The low priority class will have a backoff time greater than the high priority class. The table illustrates how to combine backoff algorithms and IFS defines four priorities from 0 to 3. ? is a random variable number in the interval (0,1), and *i* means the *ith* backoff procedure.

4.1.2 Distributed fair scheduling (backoff differentiation)

An access schema called Distributed Fair Scheduling (DFS) which utilizes the ideas behind fair queuing [6] in the wireless domain is presented in [4]. It uses a backoff mechanism of 802.11 to determine which station should send first. Before transmitting a frame, the backoff process is always initiated. The backoff interval (or contention window) computed is proportional to the size of the frame to send and inversely proportional to the weight of the flow. This causes stations with low weights to generate longer backoff intervals than those with high weights, thus getting lower priority. Fairness is achieved by including the packet size in the computing of the backoff interval, causing flows with smaller packets to get to send more often. This gives flows with equal weights the same bandwidth regardless of the packet sizes used. If a collision occurs, a new backoff interval is calculated using the DCF backoff algorithm in the 802.11 standard (the backoff interval is doubled when a transmission fails).

4.1.3 BlackBurst (BB)

The main goal of BlackBurst [5] is to minimize the delay for real-time traffic. Unlike the other schemes it imposes certain requirements on the high priority stations. BlackBurst requires :

- all high priority stations, try to access the medium with equal, constant intervals, t_{sch}, and
- the ability to jam the medium for a period of time.

When a high priority station wants to send a frame, it senses the medium to see if it has been idle for a PIFS and then sends its frame. If the medium is busy, the stations waits for the medium to be idle for a PIFS and then enters a black burst contention period. The station sends a so called *black burst* to jam the channel. The length of the black burst is determined by the time the station has waited to access the medium, and is calculated as the number of *black slots.* After transmitting the black burst, the station listens to the medium for a short period of time (less than a black slot) to see if some other station is sending a longer black burst which would imply that the other station has waited longer, thus should access the medium first. If the medium is idle, the station will then send its frame, otherwise it will wait until the medium becomes idle again and enter an other black burst contention period.

By using slotted time, and imposing a minimum frame size on real-time frames, it can be guaranteed that each black burst contention period will yield a unique winner [5]. After the successful transmission of a frame, the station schedules the next transmission attempt t_{sch} seconds in the future. This has the nice effect that real-time flows will synchronise and share the medium in a TDM fashion [5]. This means that unless some low priority traffic comes and disturbs the order, very little blackbursting will have to be done once the stations have synchronised. On the other hand low priority stations, use the CSMA/CA access method defined in 802.11 standard [1].

4.1.4 DIFS differentiation

Priority can also assigned by varying the DIFS for differentiation[7][11]. In IEEE 802.11 ACK frames get higher priority than RTS packets, simply by waiting SIFS which is shorter than DIFS (for RTS). This idea can be extended to introduce priorities for data frames (in the basic scheme) and for RTS frames (for RTS/CTS exchange). In this approach each priority level is given a different DIFS, like $DIFS_i$ where ($DIFS_i \leq DIFS_{i+1}$). So the station having priority *i* will wait $DIFS_i$ idle period before transmitting.

In order to avoid same priority frames collision, the backoff mechanism is maintained in a way that the maximum contention window size added to $DIFS_i$ is ($DIFS_{i-1} - DIFS_i$) The figure 8 shows three priority classes.

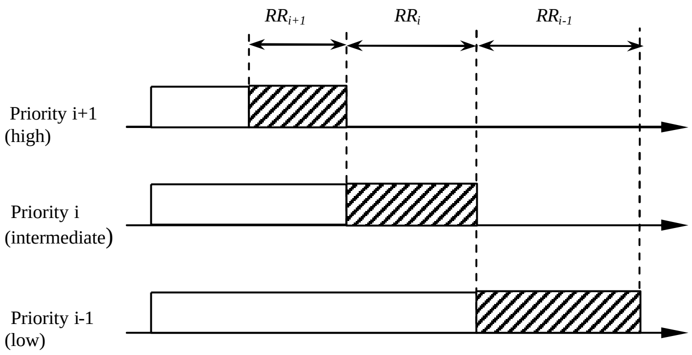

Figure 8. Including priority in DIFS

This ensures that no station with priority *i+1* has queued frames when station with priority *i* starts transmission. Low priority traffic will suffer as long as there will be high priority frames queued. It could also be the case that the maximum random range (RR_i) after $DIFS_i$ can be made greater than $DIFS_{i-1} - DIFS_i$ so the previous rule become less severe. In this case, a frame which has failed to access the medium at the first attempt will have its priority reduced after consecutive attempts, depending on the DIFSs and RRs values. This technique may be useful for realtime applications, where we have more constraints on delays than on frame drops.

4.1.5 CW$_{min}$ differentiation

The main motivation behind this CW_{min} differentiation is that with a small number of stations contending to access the channel, CW values are at their minimum value (CW_{min}) most of the time. Therefore, a backoff differentiation mechanism won't be applied correctly as the CWs are rarely increasing, and high CWs are rarely used. This leads to define differentiation on the most utilised CWs, i.e., CW_{min} [8].

In [10] a fully distributed service quality estimation, radio monitoring and admission control approach are proposed in order to support service differentiation. Two distributed estimation algorithms are defined. The virtual MAC (*VMAC*) algorithm passively monitors the radio channel and estimates key MAC level statistics such as delay, delay variation, packet collision and packet loss. A virtual source (VS) algorithm utilizes the VMAC algorithm to estimate application level service quality. The VS allows application parameters to be tuned in response to dynamic channel conditions based on "virtual delay curves". In order to provide service differentiation, both CW_{min} and CW_{max} differentiation is introduced, such as $CW_{\min}^{highprio} \leq CW_{\min}^{lowprio}$ and $CW_{\max}^{highprio} \leq CW_{\max}^{lowprio}$. Results [10] show that when the distributed virtual algorithms are applied to the admission control of the radio channel, then a globally stable state can be maintained without the need for complex centralized radio resource management.

4.1.6 Maximum frame length differentiation

This differentiation mechanism is rather simple. Different priority stations are allowed to transmit frames with different maximum frame sizes. The station with higher priority can send larger frames than a station with a lower priority. One of the following scenarios occurs:

- Drop packets that exceed the maximum frame length assigned to a given station, or
- Fragment packets that exceed the maximum frame length. This mechanism is already used in 802.11 to increase transmission reliability, and it can be used for differentiation.

This mechanism showed good differentiated results for UDP and TCP flows [7]. The major drawback of this approach is that UDP packet size is always controlled by the application, so differentiation is not really controlled by MAC sub-layer.

In the following section we present the upcoming IEEE 802.11e standard and its main new defined MAC functions.

4.2 Upcoming IEEE 802.11e QoS enhancement standard

In order to support QoS, there are some priority schemes currently under discussion [9]. The IEEE 802.11 Task Group E is currently defining enhancements to the 802.11 MAC access methods (DCF and PCF) to support QoS, providing the classes of service, enhanced security and authentication mechanism. These enhancements are defined in 802.11e which introduces a new access method called Hybrid Coordination Function (HCF). In the following paragraphs of this section, we will describe firstly this new access method of IEEE 802.11e draft D2.0 [9], secondly we will present in detail its two basis methods, the enhanced DCF (EDCF) and the HCF controlled access method.

4.2.1 Hybrid Coordination Function (HCF)

In order to support both IntServ and DiffServ QoS approaches in 802.11, TGe has defined a new mechanism called HCF. This mechanism is backwardly compatible with legacy DCF and PCF. It has both polling based and contention based channel access mechanisms in a single channel access protocol. HCF consists of two access methods, **EDCF** and **HCF** controlled channel access mechanisms. In HCF there may still be two phases of operations within the superframes, the CFP and CP, which alternate over time. The EDCF is used in the CP only, while the HCF controlled channel access method is use in both phases, which makes this new coordination function hybrid.

Three QoS levels are supported in HCF and a certain QoS classification is given in [13]. The table 2 illustrates these different QoS levels and their associated scheduling policy.

QoS levels	Channel access mechanism	Scheduling policy
Level 3	HCF (EDCF and HCF controlled channel access mechanism)	parametrized
Level 2	HCF (EDCF and HCF controlled channel access mechanism)	prioritized
Level 1	HCF (EDCF only)	prioritized
Level 0	DCF, PCF	none

Table 2. QoS levels in the HCF

A station (STA) that implements both the QoS facility and the HCF function is denoted by QSTA (QoS station) in IEEE 802.11e [9]. A Hybrid Coordinator (HC) is defined as a centralized controller in one QBSS (QoS Basic Service Set), which implements the frame exchange sequences and MSDU handling rules defined by the HCF. The HC, which is collocated within the QAP (QoS Access Point), operates during both the Contention Free Period (CFP) and the CP (Contention Period). The HC uses the point coordination’s higher priority of access to the medium to initiate frame exchange sequences and to allocate Transmission Opportunities (TXOP) to QSTAs as to provide Controlled Access Periods (CAPs) to transfer QoS data. TXOP may be allocated at appropriate times during both the CFP and CP, in order to meet predifined delivery priority, service rate, delay and/or jitter requirements of particular traffic streams. The HC may initiate Controlled Contention Intervals (CCIs) during which contention occurs only between QSTAs which need to request new TXOPs . The HCF protects the transmissions during each CAP using the virtual carrier sense mechanism (i.e., NAV) which provides improved protection of the CFP.

One of the crucial feature of 802.11e MAC is the TXOP. A TXOP is defined as the interval of time when a particular QSTA has the right to initiate transmissions onto the wireless medium. A TXOP is denoted by a starting time and a maximum duration. During the contention period (CP), each TXOP begins either when the medium is determined to be available under the EDCF rules (EDCF-TXOP) or when the QSTA receives a special frame form the HC (polled-TXOP). The duration of an EDCF-TXOP is limited by a QBSS-wide TXOP limit distributed in the beacon frames, while the duration of a polled-TXOP is specified in the header of the QoS (+)CF-Poll frame. During the contention-free period (CFP) the starting time and maximum duration of each TXOP is specified by the HC using the QoS (+) CF-Poll function. Within the limits of the duration of a TXOP, decisions regarding which MPDU (MAC Protocol Data Unit) that can be transmitted are made locally by each QSTA. By this way, internal collisions with one QSTA are avoided.

It is to be noted that in 802.11, PCF is an optional coordination function, while in 802.11e [9] HCF is a mandatory coordination function.

4.2.2 Enhanced Distributed Coordination Function (EDCF)

EDCF is a contention-based channel access scheme[9]. It is part of HCF but not a separate coordination function. EDCF provides differentiated service, distributed access to the wireless medium for 8 delivery priorities. EDCF access channel on each QSTA uses at most 8 prioritized output queues, one for each delivery priority, called Traffic Categories (TCs). The figure 9 shows the different Access Categories, i.e., 8 queues at MAC layer to support 8 Traffic Categories (TCs). Different from a station, a QAP should support at least 4 ACs. In EDCF, relative priorities are provisioned by configuring the time to access the channel once it is sensed idle and by changing the size of the contention window. EDCF uses the contention window to assign priority to each traffic category. Indeed, assigning a short contention window to a high priority TC ensures that in most cases, high priority TC is able to transmit a-head of low priority one. Thus, the CW_{min} and CW_{max} parameters can be set differently for different traffic categories, such as, a high priority TC with small values of CW_{min} and CW_{max}.

Arbitration IFS (AIFS)

For further differentiation, various interframe spaces (IFS) can be used by different traffic categories. Figure 10 shows the different interframe space relationship. Instead of using a DIFS, as a minimum specified idle duration time as defined in DCF, a new kind of interframe space called Arbitration Interframe Space (AIFS) is used. The AIFS for a given traffic category, i.e., AIFS[TC], should be a DIFS plus (possibly zero) some time slots. Thus, a traffic category (TC) with a small AIFS has a high priority.

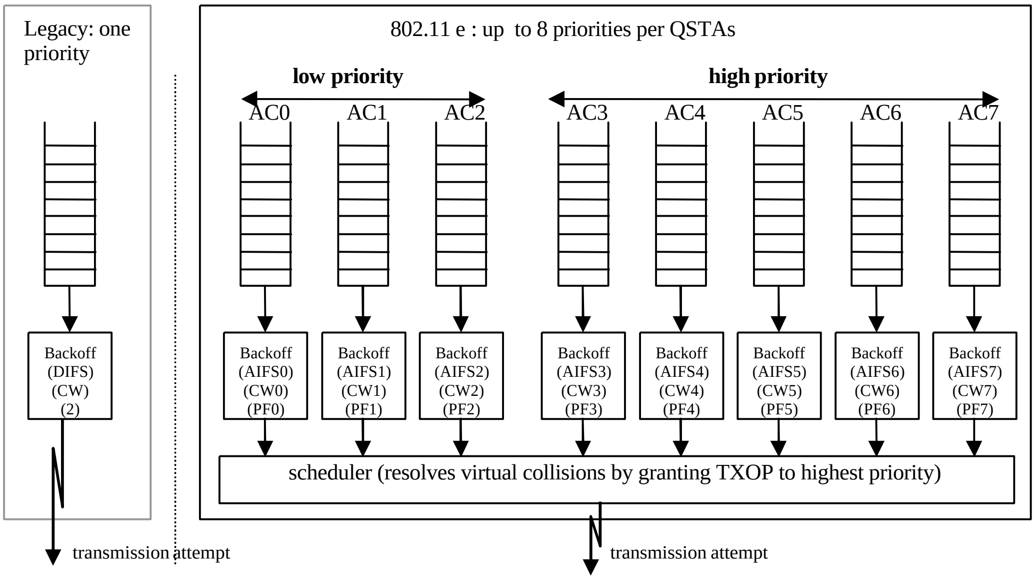

Figure 9. 802.11 EDCF vs Legacy DCF

MSDU are now delivered through multiple backoff instances within one QSTA, each backoff instance parametrized with TC-specific parameters. In the CP, each traffic category (TC) within the station, behaves like a virtual station. It contends for a TXOP and independently starts a backoff after detecting the channel being idle for at least an AIFS. The backoff time calculation uses the DCF method, but draws from different intervals and replicates the contention window state for each traffic category TC_i (or access category AC_i) as follows :

$$Backoff\_time[TC_i] = Rand(1, CW[TC_i]) \times aSlotTime$$

Where $CW[TC_i]$ is the associated contention window for traffic category TC_i, *Rand* is a random uniform distribution function over the interval $[1, CW(TC_i)+1]$, and aSlotTime is the corresponding value of the named PHY characteristics. The minimum size of the CW ($CW_{min}[TC_i]$) is another parameter dependent on the TC. Priority over legacy stations (i.e., using legacy 802.11) is provided by setting $CWmin[TC] < 15$ (in case of 802.11a PHY) and AIFS=DIFS [3].

As in legacy DCF, when the medium is sensed busy before the (backoff) counter reaches zero, the backoff has to wait for the medium being idle for AIFS again, before continuing to count down the counter. A big difference from the legacy DCF is that when the medium is determined idle for the period of AIFS, the backoff counter is reduced by one beginning the last slot interval of the AIFS period. While, with legacy DCF the backoff counter is reduced by one beginning the first slot time after the DIFS period. After any unsuccessful transmission attempt a new contention window is calculated with the help of the persistence factor PF[TC_i] and another uniformly distributed backoff counter out of this new, enlarged, CW is drawn, to reduce the probability of a new collision. Whereas in legacy 802.11, CW is always doubled after any unsuccessful transmission (i.e., equivalent to a PF=2). EDCF, uses the PF to increase the CW different for each TC_i:

$$newCW\left[TC_i\right] = ((oldCW\left[TC_i\right]+1)\times PF)-1$$

The CW never exceeds the parameter CWmax[TC_i], which is the maximum possible value for contention windows associated with the traffic category TC_i.

If the bacoff counter of two or more parallel TCs in one station reaches zero at the same time, a scheduler inside the station avoids the virtual collision by granting the TXOP to the higher priority Traffic Category (TC) or Access Category (AC), as illustrated in figure 9. At the same time, the lowest priority TCs behave as if there is an external collision on the wireless medium. However, this collision behavior does not include setting of retry counters related to the MPDU involved in collision. Since, EDCF can only resolve the possibility of internal collisions, there may still a possibility of external collisions related to the transmitted frame (they could collide at the wireless medium with frames transmitted by other stations).

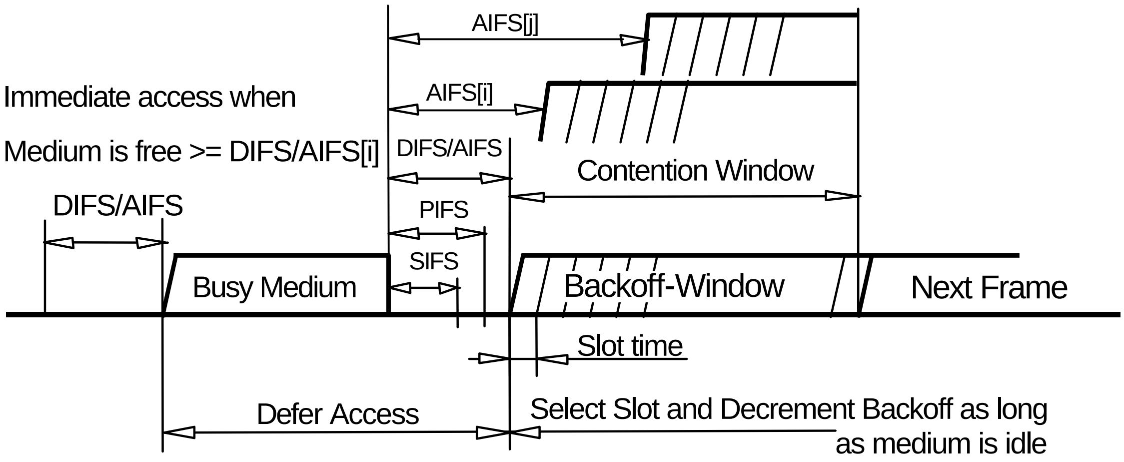

Figure 10. Some IFS relationship

In order to enhance the performance and achieve better medium utilization, packet bursting such as EDCF bursting and Contention Free Burst (CFB) can be used in 802.11e. Indeed, once a station has gained access to the medium, it can be allowed to transmit more than one frame without contending for the medium again. Indeed, after getting access to the medium, the station is allowed to send as many frames it wishes as long as the total access time does not exceed a certain limit, e.g., TxOpLimit. In order to ensure that no other station interrupts the frame bursting, a shorter IFS (i.e, SIFS) is used between frames. If a collision occurs the packet bursting is terminated. Thus, EDCF bursting can reduce the network overhead and increase throughput by multiple transmissions using SIFS and burst acknowledgement. EDCF bursting can also provide better fairness among the same priority Traffic Categories (TCs), since it is independent from the frame size. However, EDCF bursting may increase the delay jitter, so TxOpLimit should not be longer than the time required for the transmission of the largest data frame. As in EDCF, frame bursting is also provided in contention-free period during the HCF controlled channel access mode (CFB) and is presented in the following section.

4.2.3 HCF controlled channel access

The HCF controlled channel access uses a hybrid coordinator (HC) which manages the allocation of the wireless medium data transfer bandwidth. The HC is typically collocated with the QoS enhanced access point (QAP) of a QoS basic service set (QBSS). It has higher priority access than wireless stations (WSTAs), in order to allocate transmission opportunities (TXOPs) to these WSTAs (i.e., the HC can access the wireless medium only after detecting the channel is idle for PIFS, which is shorter than DIFS, thus priority over DCF traffic, and also shorter than any AIFS, priority over EDCF traffic). HC traffic delivery and TXOP allocation may be scheduled during both the CP and CFP in order to meet the QoS requirements of particular traffic categories (TCs) or traffic streams (TSs). Moreover, the HC transfers data to QSTAs during the CFP based on the amounts of queued traffic belonging to different TCs or TSs.

In HCF controlled channel access mode, Controlled Access Period (CAPs) are defined as several intervals within one CP when a short bursts of frames are transmitted using polling-based controlled channel access mechanisms. But, during the rest of the CP frames are transmitted using EDCF contention-based rules. The figure 11 shows the relationship between CFP, CP, and CAPs within one 802.11e superframe [9].

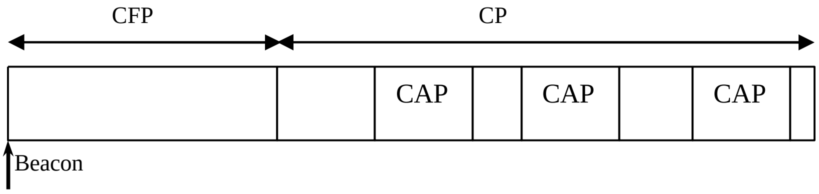

Figure 11. Relationship of CFP, CP, CAP within one superframe

CAPs may also include controlled contention (CC) periods which allow wireless stations to request the allocation of polled TXOPs without having to contend with (E)DCF traffic. These requests serve to initiate periodic polled TXOPs to handle periodic traffic under a particular Traffic Specification (TSPEC), or to handle a traffic burst or to create an initial TXOP for a new QSTA (or newly active-TS). Controlled contention is a way for the HC to learn which stations needs to be polled, at which time and for which duration. By introducing controlled contention (CC) in CAP, the HCF channel mode access can provide Guaranteed Services with a much higher probability than EDCF, specially under heavy load.

Each instance of CC occurs during the controlled contention interval (CCI) which is started when the HC sends a specific CC control frame. The CC control frame forces legacy stations to set their NAVs until the end of the CCI, thus improving protection of the CCI duration. The CC frame generated by the HC includes a priority Mask, the duration of each Controlled Contention Opportunity (CCOP), and the number of CCOPs within the CCI (NCCOP). The priority mask, is a filtering mask containing the Traffic Categories (TC) in which Reservation Requests (RR) may be placed. The figure 12 [9] shows the CC frame and the controlled contention interval (CCI) as well as the RR placed in each CCOP.

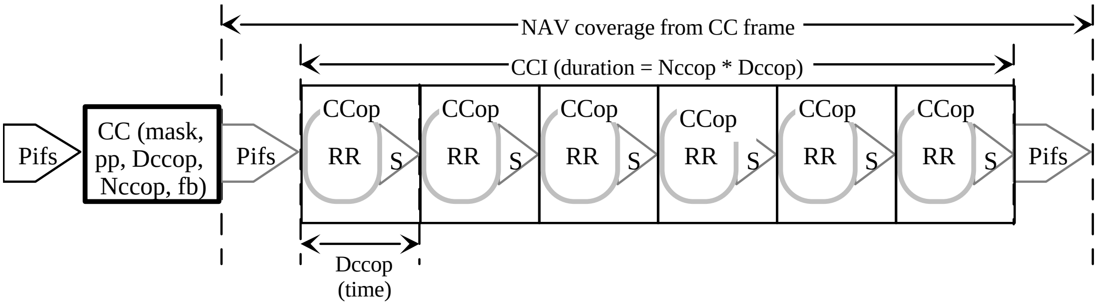

Figure 12. Controlled contention interval and different IFS (CC/RR frames)

Upon receipt of a control frame CC, each QSTA which has an RR to send for a traffic category (TC) matching the priority mask, selects a special CCOP by using a random access [9] protocol (CSMA/CA-like) to send its RR (after entering a special backoff time parameterized by Nccop and Dccop). Note that, since a random access protocol is used for RR transmission, RR frames may collide; in this case no retransmission procedure is performed. Successful reception of the request reservation frames are acknowledged by the HC to the different QSTAs in the next CCI sequence by using the feedback field (fb), i.e., fb field in the CC frame as shown in the figure 12. This mechanism enables fast collision resolution, so that the requesting QSTAs can detect collisions during the next CCI. Figure 13 illustrates an example of 802.11e superframe (also called beacon interval).

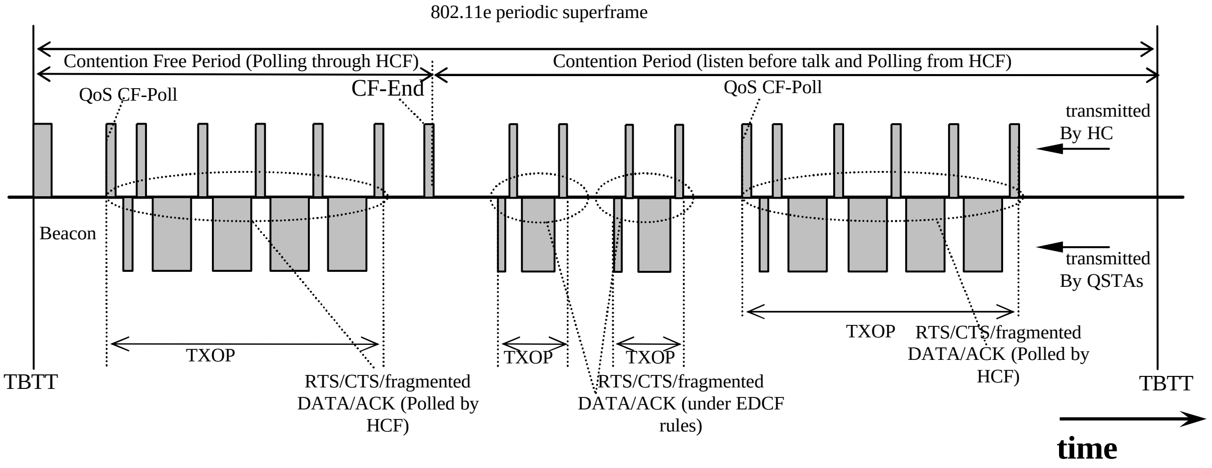

Figure 13. A typical 802.11e superframe

A superframe consists of two periods, the contention free period (CFP) and the contention period (CP). During the CFP, the starting time and maximum duration of each TXOP is specified by the HC using QoS CF-Poll frames. QSTAs will not attempt to get access to the wireless medium on its own, so only the HC can grant TXOP to these stations. The CFP ends after the time announced in the beacon frame or by an explicit frame, CF-End, from the HC. During CP, each TXOP begins either when the medium is determined to be available under the EDCF rules, i.e., after AIFS plus backoff time, or when the QSTA receives a QoS CF-Poll from the HC. The QoS CF-Poll from the HC can be sent after a PIFS idle period without any backoff (high priority WM access for the HC). Therefore the HC can issue TXOPs in the CP using its prioritized medium access.

In order to achieve a better medium utilization, a Contention Free Burst (CFB) can take place during the CFP. CFB allow a sequence of MPDUs (more than one MPDU) to be transmitted during a TXOP, separated by a SIFS period. The MPDU within a burst is acknowledged by a BurstAck MPDU, requeted by a BurstAck request sent from the originating WSTAs. This “request”/ “response” mechanism gives the recipient time to perform any necessary FEC decoding.

The burst is started by receiving a QoS CF-Poll frame from the HC. It does not require a traffic specification (TSPEC) and it applies to FEC and non-FEC use. The duration values of the burst data MPDUS and any burst ACK exchange should fit within a single TXOP (i.e., polled TXOP duration specified in the QoS CF-Poll frame). The main purpose of CFB is that enables the HC to give up medium control (to possibly another QBSS) and defer control until new frames of other CFB are available. By this way the HC can relinquish medium control to other QBSSs in the same area to solve the problem of BSS overlap [13].

4.2.4 Performance of 802.11e

Simulation based evaluations of EDCF, HCF under various load conditions, has been done in the context of related work [3][7][12][13][14]. Performance analysis [13], shows that EDCF scheme can support better QoS than DCF and PCF during low medium load conditions. However, the EDCF based ad-hoc network saturates and throughput decreases when the load increases. Thus, it is very difficult to find the optimal EDCF parameters that can give the best performance, since the parameters are static and can not adapt to the traffic load rate. At the same time HCF controlled channel access uses a QoS-aware scheme for different queues in each QSTA, and performs better than EDCF. However, no admission control policy is specified for HCF and without a such policy, it works well only until a certain traffic (i.e., QoS load upbound) load and fails when load exceeds the QoS-load upbound (results from none public technical reports related to 802.11e evaluation [21]).

In the context of 802.11e evaluation, we are planning in the future work to more study and evaluate the different QoS schemes, i.e., EDCF and HCF, in order to better apprehend their behavior under different wireless channel conditions (see next section).

5 Open issues and future work

There are still many open issues and research work to be done in order to validate the proposed QoS mechanisms for IEEE 802.11e standard [13]. Indeed, the proposed standard is still unapproved, unstable and thus needs to be tested more. There are still many open issues among them, we cite :

- Interference problem between QBSSs : still not studied, i.e., robustness of the polling scheme under HCF as compared to EDCF when there is interference (from other AP, from station, or from a hidden node).
- Optimal EDCF : decreased performance in high load environment, research is carried out to adapt the parameters to the traffic load and channel condition for ad-hoc EDCF mode.
- HCF validation : scheduling algorithms, for guaranteed service, and trade-off between channel efficiency, priority and fairness. Need for an admission control policy.
- Evaluate the efficiency and performance of EDCF packet bursting and contention-free burst (CFB).
- Cooperation between MAC level FEC and other higher layers error control protocols.
- Performance of 802.11e with different QoS requirements and under different operational scenarios.

In the following we present the main tasks that we intend to carry out for future work:

- Simulation study and analysis of 802.11e QoS mechanisms (EDCF) : In order to better evaluate EDCF QoS scheme, it will be interesting to design some more relevant scenarios (different QSTAs with different TS QoS requirements, i.e., several audio, video and data traffic sources) and tune EDCF parameters in order to have the best performance for EDCF. It is also interesting to adapt EDCF parameters (Slow CW decrease approach) to the traffic load and channel conditions especially in case of high load in QBSS (e.g., AEDCF; adaptive EDCF approach is an example [12], but did not solve all the problems). Use of EDCF NS code implementation of INRIA, available at : http://www-sop.inria.fr/planete/qni/Research.html. The goal is to analyse EDCF/ad-hoc EDCF via simulations and to propose if necessary extensions. Consider asymmetric traffic in load and size of frames (uplink/downlink).
- HCF analysis and simulation : design and evaluate various HCF scheduling algorithms which can be used by the HC in order to meet the QoS requirements for the different Traffic Categories (TCs). Especially, taking into account CBR traffic requirements in the scheduling parameters/algorithms (e.g., determine the limit of simultaneous QSTAs/TCs with CBR traffic that can be admitted inside a QBSS), via simulation and comparison with EDCF and related work. Consider also asymmetric traffic in load and size of packets. The HCF implementation of Stanford & Atheros Communications, will be used and extended [16].
- EDCF Implementation (goal is to provide QoS support with EDCF): Use of an Open source DCF code, enhance and extend it with prioritization classes (4 or up to 8 classes). Implement different back-off algorithms, local QSTA scheduler in order to resolve virtual collisions. Tests evaluation with a benchmark scenarios on one station and on different stations within the same AP. Use linux-WLAN or HostAP available implementation of the legacy DCF access method. The goal is to extend HostAP/linux-WLAN with EDCF QoS features.
- Interference problem : multiple access point scenarios create cross-cell interference particularly hurting polling scheme due to sychronization. The purpose is to study the robustness of the polling scheme under HCF as compared to EDCF when there is interference (from other AP, from station, or from a hidden node). Determine the impact of interference on the performance of HCF via simulating relevant scenarios. The goal is to propose and evaluate various HCF scheduling algorithms in order to cope with the interference problem.
- Standard tracking and contribution : following invitation from chair IEEE Task Group e (TGe), it may be interesting (but vital) to participate to the next meetings of IEEE TGe (scheduled for May or July

2003), in order to track the evolution of the 802.11e, to have all related technical materials (a lot of reports from active TGe members on different issues of 802.11e) and to participate in its validation and approval process.

- HCF modelling and Performance Evaluation. Delay, jitter, bandwidth and utilization modelling and analysis.
- FEC-based MAC error recovery : adequacy of FEC in MAC? fading error model problem in WLAN.

6 Conclusion

This report, in a first part, introduces 802.11 protocols and analyzes QoS limitations. Legacy 802.11 DCF and PCF access methods provide only a best-effort service for wireless users. A lot of research activities have been carried out to improve the performance of 802.11 WLAN. We provided a survey of the proposed QoS techniques and enhancements for 802.11 MAC layer protocols. Main QoS techniques enhance the DCF access method to provide differentiated services by assigning high priorities to certain stations. The main features of the new access methods defined for the upcoming 802.11e QoS-based standard are presented. EDCF and HCF constitute the two main access methods enabling to provide, respectively, differentiated services and guaranteed services as basis for QoS support in WLAN. Certain drawbacks and performance limitations of IEEE 802.11e are also addressed.

It would be interesting to note that many open issues and research work has to be done in order to validate the proposed QoS mechanisms for IEEE 802.11e standard, namely EDCF and HCF. Indeed, 802.11e is still unapproved, unstable and thus needs to be more tested. Therefore, the maturity path of IEEE 802.11e is still long and its validation process is ongoing, thus it may be a chance for us to contribute to the upcoming standard.

7 Acknowledgement

This work is supported by the Hes-so through the CCTI (Competence Centre for Information Technology) fund under grant number CCTI 02-10. It is also partly supported by Swisscom Innovations. We would like to thank Swisscom Innovations' Broadband Networks team, especially Beat Perny and Jan Linder for their valuable remarks which helped to improve this report.

8 References

[1] IEEE WG, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE 802.11 Standard, 1999.

[2] D-J. Deng and R-S. Chang. A priority scheme for IEEE 802.11 DCF access method. IEICE Transactions on Communications, E82-B(1), January 1999.

[3] Mangold, S. and Choi, S. and May, P. and Klein, O. and Hiertz, G. and Stibor, L. "IEEE 802.11e Wireless LAN for Quality of Service", in Proc. European Wireless '02, Florence, Italy, February, 2002.

[4] N. H. Vaidya, P. Bahl, and S. Gupta. Distributed fair scheduling in wireless LAN. In Sixth Annual International Conference on Mobile Computing and Networking, Boston, August 2000.

[5] J. L. Sobrino and A. S. Krishnakumar. Quality of Service in ad-hoc carrier sense multiple access networks. IEEE Journal on Selected Areas in Communications, 17(8): 1353-1368, August 1999.

[6] S. J. Golestani. A self-clocked fair queuing scheme for broadband applications. In Proceedings of IEEE INFOCOM, 1994.

[7] Imad Aad and Claude Castellucia. Differentiation mechanisms for IEEE 802.11. In Proceedings of IEEE Infocom 2001, Anchorage, Alaska, April 2001.

[8] I. Aad and C. Castelluccia. Remarks on per-flow differentiation in IEEE 802.11. European Wireless 2002, Florence, Italy, February 25th-28th, 2002

[9] IEEE WG, Draft Supplement to Standard for Telecommunications and Information Exchange Between Systems-LAN/MAN specific requirements –Part 11: Wireless Medium Access Control (MAC) and Physical Layer () specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), IEEE 802.11e /Draft D2.0, November 2001.

[10] A. Veres, A.T. Campbell, M. Barry, and L-H. Sun. Supporting service differentiation in wireless packet networks using distributed control. IEEE Journal of Selected Areas in Communications (JSAC), Special Issue on Mobility and Resource Management in Next-Generation Wireless Systems, vol. 19, No. 10, pp. 2094-2104, October 2001.

[11] Drabu, Y. A Survey of QoS techniques in 802.11. Available at: http://trident.mcs.kent.edu/~ydrabu/research/wmac/mac.pdf

[12] Romdhani L., Ni. Q., And Turletti T. AEDCF: Enhanced Service Differentiation for IEEE 802.11 Wireless ad-hoc. INRIA report No 4544, Sept. 2000.

[13] Qiang Ni, Lamia Romdhani, Thierry Turletti, and Imad Aad. QoS issues and Enhancements for IEEE 802.11 Wireless LAN. INRIA report No 4612, November 2002.

[14] A. Lindegren, A. Almquist, and O. Schelén. Quality of service schemes for IEEE 802.11, a Simulation Study. IWQoS 2001.

[15] Ritabrata, R. Traffic and QoS performance of an Infrastructure CSMA/CA Network with Multiple Service Classes. Report of the university of Rutgers, state University of NJ. Available at : http://www.winlab.rutgers.edu/~rito/

[16] Garg P., Doshi R., Greene R., Baker M., Malek M., Cheng X., Using 802.11e MAC for QoS over Wireless. Report of the University of Stanford. Available at : http://milliways.stanford.edu/~radoshi/cs444n/

[17] Mangold, S. QoS support as Utility for Co-existing Wireless LANs. In Proceedings of the International Workshop on IP Based Cellular Networks, IPCN, Paris, France, April 2002.

[18] Mangold, S. Fair Resource Sharing Between Overlapping Basic Service Set. In Proceedings of the PIMRC 2002, pp. 166-171, Lisbon, Portugal, September 2002.

[19] Tourrilhes, J. Linux and WLAN. Available at : http://www.hpl.hp.com/personal/Jean_Tourrilhes/Linux/

[20] S. Gats, Matthew. 802.11 Wireless Networks, the Definitive Guide. 2002 O'Reilly & Associates.

[21] IEEE 802.11 Task Groups Home page. Available at http://grouper.ieee.org/groups/802/11

[22] Perkins, D.D., and D. Huges, H. A Survey on Quality of Service Support for Mobile ad-hoc Networks. Wireless Communications and Mobile Computing. 2002; 2:503-513

[23] Banchs, A. and Perez, X. Providing Throughput Guarantees in IEEE 802.11 Wireless LAN. WCNC2002, Vol. 1, pp. 130-138.

[24] Tanenbaum, AS. Computer Networks, 3rd Edition, Prentice Hall Edition, New Jersey, 1996.

[25] T.Gandolfo, P. Trade-Off Analysis (802.11e versus 802.15.3 QoS mechanism), XtremeSpectrum technical report.

[26] Chandra, A., Gummala, V. and Limb, J. Wireless Medium Access Control Protocols. IEEE Communications, Surveys & Tutiorials. Sept. 2000.

[27] H. Schiller, J. Mobile Communications, Addison-Wesley, 2000.

9 Annexe 1: Relevant Access Point (AP) Products Supporting QoS in WLANss

Main classes of Wireless Access Points (AP)

This annex presents different types of available access points (APs) and organise them into different classes which are relevant to our project (i.e, QoS for 802.11 WLAN). This annex do not explore all available APs in the market. Hence, there may be other relevant APs that we could forget to report here.

A. Traditional 802.11 AP products

- *Proxim ORINONO* (formely Lucent Wavelan, Agere) : support of 802.11b standard CSMA/CA access method. No QoS support at the air link.
- *Symbol AP*: support of mobile-IP for seamless roaming. Compliant with 802.11b standard. No QoS support. Use of 802.1p to provide QoS at layer 2 Ethernet frames.
- *Cisco Aironet* : networking functions, modular and manageable solution. Limited support for QoS at the air link. Only a QoS support for packets travelling from the access point *over the Ethernet.* The access point supports QoS for voice over IP (VoIP) telephones and downlink prioritized channel access for streaming audio and video traffic. Capability to upgrade to future 802.11e standard.

B. Emerging 802.11 AP products with QoS support

- *Zaye Tech (WA222M AP)* : is an advanced wireless multimedia Access Point/Bridge that combines innovative 802.11e technology with multimedia QoS support. ZayeTech AP is based on Cirrus Logic WhiteCap2 network protocol which is the wireless LAN solution capable of transmitting high quality audio and video throughout the home. It addresses specifically the requirements for wireless home networks, including multimedia support, interference immunity, extended range and ease of use.
- *Cirrus Logic (Bodega AP CS22250) :* main product providing QoS for complete multimedia support, including contention free support for deterministic behavior and predictable latency, supports for multiple streams and dynamic allocation of resources. Support of up to 24 WiFi nodes and 10 multimedia nodes per AP/Bridge. Today, PCF/HCF access methods are rarely implemented and this is due to some hardware constraints and partly to MAC protocols immaturity. Cirrus Logic through this product claims to implement the main relevant access methods for multimedia support.
- *AtherosCcommunications* : support of QoS as defined in 802.11e draft. Context of 802.11a networks. Have two Ethernet interfaces capable to connect to both an Ethernet LAN and an existing 802.11b WLAN access point. Use of AES (advanced encryption standard) for security concerns.

C. New technology for Mobile Broadband Wireless Access Networks (with high QoS support)

- *Flarion RadioRouter* : Support of QoS for multimedia communications. Context of mobile broadband wireless access networks. Provides 1.5 Mbps per sector and up to 3 Mbps for a peak rate per user. Flash OFDM technology is the heart of this new technology and is under way for standardisation (actually is not a standard, but an IEEE 802 has been set up). With Flash OFDM there is no interference between users within the same cell and sector. Furthermore, there is minimal interference among users from neighboring cells (spread spectrum approach)

We note that all AP are with Ethernet interface ,i.e., IEEE 802.3 10BaseT, for infrastructure networks topology. Table 1 gives a summary for the main features that characterize these classes of APs and hence could provide a comparison basis.

Abbreviations- related to products' features

PCF	Point Coordination Function
HCF	Hybrid Coordination Function
DSSS	Direct Sequence Spread Spectrum
OFDM	Orthogonal Frequency Division Multiplex
WEP	Wired Equivalent Privacy
AES	Advanced Encryption Standard
TKIP	Temporal Key Integrity Protocol
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
MBWA	Mobile Broadband Wireless Access

Feature **Product**	Frequency Band	Network Standard	Network Architecture	Modulation	Security	Quality of Service (QoS)	Media Access Type	Data rates	Misc. Information Home page
Proxim ORiNOCO AP	2.4GHz up to 5GHz	Upgradeabele-from 802.11a, 802.11b, 802.11g	Infrastructure	DSSS OFDM	802.1x support auto key management	No mention of QoS support	CSMA/CA	11Mbps and 54Mbps	www.orinocowireless.com ORiNOCO AP-600
Cisco Aironet 1200 series	2.4 GHz up to 5GHz	802.11b and 802.11a	Infrastrcuture And star topology	DSSS OFDM	DSSS OFDM	Limited QoS. Support of QoS for 802.1p	CSMA/CA	Up to 11Mbps 802.11b up to 54Mbps	http://www.cisco.com/warp/public/44/jump/wireless.shtml Networking functions and manageable solution
Symbol AP	2.4GHz to 2.5GHz	802.11b	Infrastructure Ad-hoc	DSSS	WEP	No support for QoS at the Air link. 802.11p	CSMA/CA	Up to 11Mbps	www.symbol.com Roaming virtually instantaneous Support of Mobile IP
CirrusLogic Bodega AP	2.4GHz	802.11b 802.11e - type	Infrastructure	DSSS	Device authentication and 40 bit encryption	Strong support of QoS	CSMA/CA HCF	Up to 11Mbps	www.cirruslogic.com Parametrized QoS (bandwidht, latency and jitter) Multimedia Support
ZayeTech	ISM band 2.4GHz (2.400-2.497)	802.11b 802.11e-type	Infrastructure Ad-hoc	DSSS	Data Encryption 64/128 bit WEP	Support of QoS 802.11 e-type	Contention-free access and CSMA/CA	Up to 11 Mbps	www.zayetech.com Zaye Tech WA222M
Atheros Comm.	4.900-5.850 GHz	802.11a	Infrastructure /Ad-hoc	OFDM	Encryption; WEP, AES, TKIP Authentication : 802.1x	QoS support compliant with 802.11e draft	CSMA/CA	6-54 Mbps Atheros Turbo : 12-128 Mbps	www.atheros.com Entreprise Quality functionality
Flarion Flash-OFDM	Licensed radio frequence up to 3.5 GHz	New : Under standardisation.	Infrastructure	Flash-OFDM technology		Strong support of QoS : Air link QoS	MBWA	LAN-like experience to data users : 1.5 Mbps/sector	www.flarioncom Flarion Radiorouter can interface with GSM/UMTS Support of Mobile-IP

Table 1. Summary of main Access Point Features