Available online at www.sciencedirect.com

ScienceDirect

Computer Communications 29 (2006) 2621–2633

computer communications

www.elsevier.com/locate/comcom

A performance study on service integration in IEEE 802.11E wireless LANs

Frank Roijers [a,*], Hans van den Berg [a,b], Xiang Fan [a], Maria Fleuren [a]

[a] *TNO Information and Communication Technology, Delft, The Netherlands*
[b] *University of Twente, Department of Design and Analysis of Commmuncation Systems, Enschede, The Netherlands*

Available online 20 March 2006

Abstract

Several studies in literature have investigated the performance of the proposed IEEE 802.11E standard for QOS differentiation in WLAN, but most of them are limited both with respect to the range of the parameter settings and the considered traffic scenarios. The aim of the present study is twofold. First, we systematically investigate the differentiating capabilities of QOS mechanisms. Second, we investigate how well the QOS mechanisms are able to support different types of services under realistic traffic conditions. In particular, we investigate flow-level performance characteristics (e.g., file transfer times) in the situation that the number of active stations varies dynamically in time.

Keywords: Wireless LAN; IEEE 802.11E; EDCA; Quality of service; Flow level performance

1. Introduction

A major drawback of existing versions of the IEEE 802.11 WLAN standards, notably the widely used IEEE 802.11B [8] version, is that they are not capable of providing any service guarantees. The most widely deployed IEEE 802.11B MAC protocol, the so-called Distributed Coordination Function (DCF), is a random access scheme based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). Current research and standardization efforts are aiming at enhancements of the DCF MAC protocol enabling the support of multi-media applications with stringent QOS requirements. In particular, the Enhanced Distributed Channel Access (EDCA) of the IEEE 802.11E standard [9], which was finalized mid 2005, provides several parameters enabling QOS differentiation among the traffic originating from applications with different service characteristics. Existing studies on the QOS provisioning capabilities of IEEE 802.11E are limited both with respect to the range of the parameter settings and the assumptions about the traffic generated by the WLAN stations/users. Therefore, the aim of the present study is to systematically investigate (by extensive simulations) the impact of the QOS differentiation parameters, under more realistic traffic conditions. In particular, besides considering traffic scenarios with a fixed number of persistently active stations (as assumed in most other studies) we also investigate flow-level performance characteristics (e.g., file transfer times) in the situation that the number of active stations varies dynamically in time.

1.1. Related literature

For the 802.11B version several analytical models have been developed in order to study the system's saturation throughput as a function of the number of (persistently) active stations. The most well-known model is the one developed by Bianchi [2]. It is based on a Markov chain describing the behavior of a single station attempting to send its packets. Foh and Zuckerman [5] and Litjens et al. [12] investigate the flow-level performance of

* Corresponding author. Tel.: +31 15 285 7244; fax: +31 15 285 7057.
E-mail addresses: frank.roijers@tno.nl (F. Roijers), j.l.vandenberg@telecom.tno.nl (H. Berg), XFan@euro.delta-corp.com (X. Fan), maria.fleuren@tno.nl (M. Fleuren).

0140-3664/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2006.01.034

802.11B WLAN when the number of active stations varies in time, e.g., due to the random initiation and completion of file transfers. In particular, building on Bianchi's work [2], they obtain accurate approximations for the mean flow transfer delay.

Most performance studies of the QOS-enabled 802.11E WLAN are based on simulation. Relatively few papers present an analytical approach. E.g., [17] propose extensions of the Markov chain analysis presented in [2] for the 802.11B version, in order to capture the impact of variation of the AIFS parameter (one of the EDCA parameters) on the saturation throughputs. Both analytical models yield accurate results. [3] uses a queueing system with Discriminatory Processor sharing (DPS) service discipline to model flow-level behavior; results are validated by simulations. The simulation studies usually consider more general scenarios (sometimes also capturing the impact of higher layer protocols like TCP), but a systematic study of the impact of each of the EDCA parameters on WLAN performance is lacking. In particular, [4,7,11,13] compare the 802.11B version with the 802.11E version, but only for the default parameter settings; other papers (e.g., [1,6,10,16]) consider a broader range of parameter values but only for some of the QOS differentiation parameters.

Most of the studies mentioned above assume a fixed number of persistently active stations. In some cases (see e.g.,[1,4,6,7]) the impact of adding one or two additional (persistently active) stations is studied by plotting the resulting throughputs as function of the time. However, flow-level performance studies, which take into account that the number of active stations varies dynamically in time, are not available.

1.2. Contribution

Our first contribution is a systematic evaluation of the IEEE 802.11E QOS differentiation parameters (EDCA parameters) CW_{min}, CW_{max}, AIFS, and the $TXOP_{limit}$. The differentiating capabilities are studied in a scenario with two groups of persistent stations in order to exclude user behavior and influences of higher OSI-layers. First each EDCA parameter is studied in isolation, i.e., only the parameter of interest has a different value per group, while the other parameters are set equally for both groups. Subsequently, the parameters are studied in competitive scenarios; the parameters of both groups are set according to the 802.11B standard except for two parameters which are differentiated. The main performance metrics that are investigated are the resulting throughputs per station and per group.

Our second contribution is a thorough investigation of the EDCA's capabilities to provide QOS guarantees in a more realistic scenario with a dynamically varying number of active stations. We consider three different service classes: voice, video, and (TCP controlled) data traffic. The main performance metrics, studied by simulation, are packet delay (particularly important for voice), packet loss (important for video) and data flow (file) transfer time.

1.3. Outline

The remaining of this paper is organized as follows. In Section 2, the main principles of the 802.11E MAC protocol and its mechanisms to provide QOS differentiation are explained. In Section 3, the simulation scenarios are described both for the differentiating capabilities of the EDCA parameters (assuming persistently active stations) and the flow-level study. In Section 4, the results of the simulation studies are presented and discussed. Section 5, concludes this paper.

2. IEEE 802.11E QoS enhanced wireless LAN

In this section we briefly explain the IEEE 802.11B Distributed Coordination Function and its enhancements as specified in IEEE 802.11E in order to support QOS differentiation. In this paper we only use the BASIC access mode.

2.1. IEEE 802.11B Distributed Coordination Function

Fig. 1 illustrates the principle of the BASIC access scheme. When a station wants to transmit a data packet, it first senses the medium to determine whether or not the channel is already in use by another station (*physical carrier sensing*). If the channel is sensed idle for a contiguous period of time called DIFS (Distributed InterFrame space), the considered station transmits its packet. In case the channel is

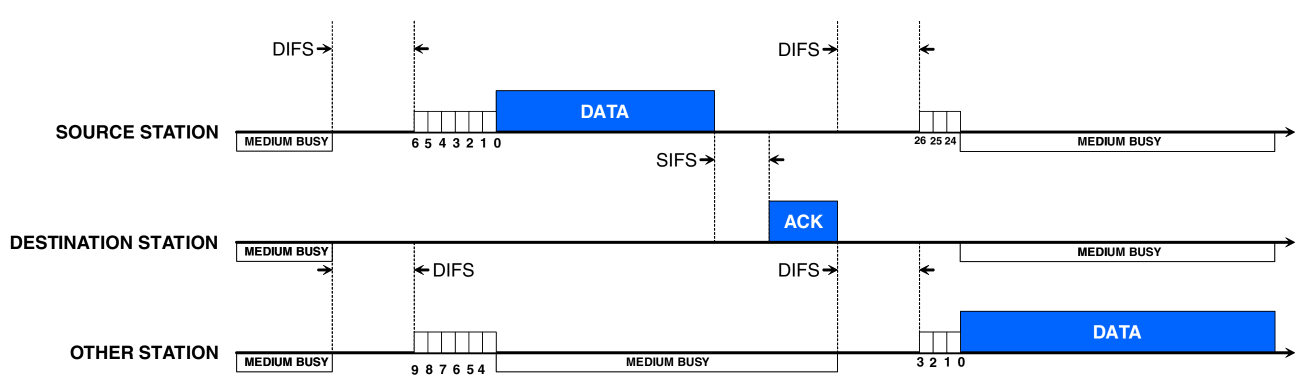

Fig. 1. Illustration of the BASIC access mode of the Distributed Coordination Function of the IEEE 802.11 MAC layer.

sensed busy, the station must wait until it becomes idle again and subsequently remains idle for a DIFS period, after which it has to wait another randomly sampled number of time slots before it is permitted to transmit its data packet. This *backoff* period is sampled from a discrete uniform distribution on $\{0,\ldots,\mathrm{CW}_r - 1\}$, with CW_r the contention after r failed packet transfer attempts (CW_0 is the initial contention window size). The backoff counter is decremented from its initially sampled value until the packet is transferred when the counter reaches zero, unless it is temporarily 'frozen' in case the channel is sensed busy before the backoff counter reaches zero. In the latter case the station continues decrementing its backoff counter once the medium is sensed idle for at least a DIFS period. It is noted that the idea behind the random backoff procedure is to reduce the probability of *collisions*, which occur either when the backoff counters of multiple stations reach zero simultaneously, or in case a so-called hidden station fails to freeze its backoff counter when it cannot sense another station's transmission. In a collision only the strongest signal among multiple concurrent transmissions has a chance of successful *capture* by the intended receiver.

If the destination station successfully captures the transmitted data packet, it responds by sending an ACK (ACKnowledgement message) after a SIFS (short InterFrame space) time period. A SIFS is shorter than a DIFS in order to give the ACK preference over data packet transmissions by other stations, while it is sufficiently long to allow the stations involved in the considered transfer to switch between transmission and reception mode. If the source station fails to receive the ACK within a predefined time-out period, the contention window size is doubled unless it has reached its maximum window size, upon which the data packet transfer is reattempted. The total number of transmission attempts is limited to $r_{\max}$. Once the data packet is successfully transferred, the contention window size is reset to CW_0 and the entire procedure is repeated to transfer subsequent data packets. If an unfortunate data packet is still not successfully transferred after $r_{\max}$ retransmissions, the MAC layer gives up. It is then up to higher-layer protocols (e.g., UDP (User Datagram Protocol) or TCP) whether the packet is discarded or once again offered to the MAC layer for transmission.

2.2. IEEE 802.11E Enhanced Distributed Channel Access

IEEE 802.11E specifies the Enhanced Distributed Coordination Access (EDCA) as the distributed contention mechanism that can provide service differentiation. Whereas an 802.11B station has only one queue for all traffic, an 802.11E station (QSTA) has multiple queues, so-called Access Categories (ACs), and traffic is mapped into one of the ACs according to its service requirements. Each AC contends for the medium using the CSMA/CA mechanism described in Section 2.1 using it own set of EDCA parameters values. These EDCA parameters are $\mathrm{CW}_{\min}$, $\mathrm{CW}_{\max}$, AIFS, and the $\mathrm{TXOP}_{\mathrm{limit}}$.

The parameters $\mathrm{CW}_{\min}$ and $\mathrm{CW}_{\max}$ have the same functionality as in the DCF. The parameter AIFS (Arbitration InterFrame space) differentiates the time that each AC has to wait before it is allowed to decrement its backoff counter after the medium has become free. In the DCF each station has to wait for a DIFS period while the duration of an AIFS is a SIFS period extended by a discrete number of time slots AIFSN, so AIFS = SIFS + AIFSN × timeslot (where AIFSN $\geqslant 2$ for QSTAs and AIFSN $\geqslant 1$ for Quality APs). The $\mathrm{TXOP}_{\mathrm{limit}}$ (Transmission Opportunity limit) is the duration of time that an AC may send after it has won the contention, so it may send multiple packets as long as the last packet is completely transmitted before the $\mathrm{TXOP}_{\mathrm{limit}}$ has passed. Fig. 2 illustrates the parameters AIFS and $\mathrm{CW}_{\min}$. Obviously, the backoff counters of multiple ACs of one station can reach zero at the same moment, which is called a *virtual collision*. Each QSTA has an internal scheduler that handles a virtual collision. The AC with the highest priority is given the Transmission Opportunity (TXOP) and may actually initiate a transmission. The ACs of lower priority are treated as if they experienced a collision, so they have to double their contention window CW and start a new contention for the medium.

3. Description of the simulation scenarios

This section describes the scenarios and parameter settings that are used in the simulation studies.

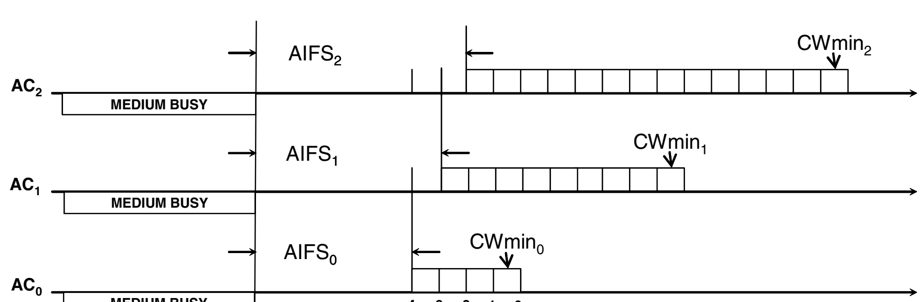

Fig. 2. QoS station with three Access Categories.

3.1. System model

We consider a single Basic service set (BSS) with stations contending for a shared radio access medium with a channel rate of $r_{WLAN} = \{1,11\}$ Mbit/s. The physical layer preamble is always transmitted at 1 Mbit/s and the rate of the MAC layer preambles is {1,2} Mbit/s. All stations are assumed to have comparable radio conditions so that a uniform channel rate can be assumed. Only the BASIC-access mode is considered in the simulations.

The simulations are performed using the Network simulator NS-2 [14] extended by the EDCA implementation of the TKN Group of the Technical University of Berlin [15]. This implementation contains the differentiation parameters explained in the previous section. Packet capture, which is the possibility that a packet with a strong signal may survive a collision, is turned off in this study.

3.2. Traffic scenarios

EDCA performance is studied for two main traffic scenarios. In Scenario 1, the impact of the EDCA QOS differentiation parameters is studied assuming persistently active traffic sources (stations). In Scenario 2, the QOS differentiation capabilities of EDCA in the case of non-persistent traffic sources (i.e., a dynamically varying number of active stations) are investigated.

3.2.1. Scenario 1: persistent traffic sources

In Scenario 1, the number of active stations remains fixed during a single simulation. Each station generates traffic in the upstream direction and is assumed to always have traffic available for transmission. The traffic consists of 1500 Byte IP/UDP packets and all data and headers are transmitted at 1 Mbit/s.

First, we study differentiation by a single parameter. Each station uses only 1 AC and stations are divided into two groups, AC_0 and AC_1. All parameters are set according to the 802.11B standard, except for AC_0 stations where one parameter is set to a more favorable value. In the simulations, the total number of stations n is increased in two different manners: in one series the number of stations is always equal for both groups (so each group has $n/2$ stations), and it the other series there are always 2 AC_0 stations and the remaining (n-2) stations are AC_1 stations.

Second, we study the differentiating capabilities by different parameters. Again stations are divided into two groups and parameters are set 802.11B settings, but now for each group a different parameter is given a more favorable value. The number of stations per group is always equal. This scenario allows us to examine which of the differentiated parameters dominates the other parameter.

In all scenarios, the throughput per station and the aggregate throughput per group are obtained. Also the throughput ratio of an AC_0 station and an AC_1 station is studied. The investigated parameter settings are (802.11B values are denoted in boldface): $CW_{min} = \{7, 15, \mathbf{31}\}$, $CW_{max} = \{31, 127, \mathbf{1023}\}$, $AIFSN = \{1, \mathbf{2}, 5\}$ and $TXOP_{limit} = \{\mathbf{0}, 0.03, 0.1\}$ s.

3.2.2. Scenario 2: dynamic user scenario (non-persistent traffic sources)

In Scenario 2, the number of active stations varies dynamically during a simulation due to e.g., the initiation and completion of speech calls or web page downloads. Three different Access categories are considered corresponding with voice over IP (VOIP, an interactive service), video-on-Demand (VOD, a streaming service), and web Browsing (an elastic data service). For each of these services we will consider below the main characteristics and modeling assumptions made in our simulations; specific modeling assumptions are summarized in Table 1.

VOIP is a real-time, interactive service and requires the end-to-end delay to be less than 150 ms. Besides the delay also the packet loss is constrained; it should be less than a few percent. In the simulations new VOIP calls are initiated according to a Poisson process and a VOIP-call is modeled by two UDP CBR streams (80 Kbit/s each).

VOD traffic is sent at a fixed rate from a video server to a user. The most important QOS-constraint for streaming video is packet loss as video-codecs are very sensitive to loss. Packet delays are less important. In the simulations, a VOD traffic stream is modeled by a UDP CBR packet stream (480 Kbit/s). The VOD calls are generated by a fixed number of users; the time between the completion of a VOD call and the initiation of a new call by a particular user is exponentially distributed.

web-Browsing is controlled by TCP. The most important QOS metric for this application type is the web page download time or, closely related, the throughput during a web page download. In the simulations web page downloads are initiated according to a Poisson process. Web pages are retrieved from a web server that is connected to the AP by a fixed link with a certain capacity and transmission delay. The capacity is chosen such that it is not a bottleneck and no packets will be lost.

Table 1
Services classes

	Voice over IP	Video on demand	Web-browsing
Flow-level parameters			
Transport protocol	UDP	UDP	TCP
Downstream traffic	CBR	CBR	TCP data
Upstream traffic	CBR	–	TCP ACKS
Arrival process	Poisson	ON–OFF	Poisson
Arrival rate	1/60	–	4
ON-time/file size distr.	exponential	exponential	exponential
Avg. ON-time	180 s	300 s	–
Avg. file size	–	–	15 KBytes
OFF-time distribution	–	exponential	–
ON–OFF ratio	–	1:4	–
Packet-level parameters			
IP packet size	200 Bytes	1500 Bytes	1500 Bytes
bit rate	80 Kbit/s	480 Kbit/s	

In Scenario 2, the WLAN operates at 11 Mbit/s. Starting with a certain mix of offered traffic (determined by the default parameter settings shown in Table 1), we study the effects of increasing the traffic load by one of the service class, while the traffic load from the other service classes remains unchanged. The performance metrics of interest are packet loss, packet delay, and delay jitter and the mean web page download time; because of lack of space we will omit in this paper the performance results for the VOD and WB service classes. In the simulations, the total number of users simultaneously present in the system is at most 50 as new users are blocked when already 50 users are present.

4. Numerical results

This section presents the simulation results using the system model parameters settings of Table 2. Section 4.1 presents the results of the Persistent Traffic Sources scenario described in Section 3.2.1 and Section 4.2 presents the results of the Dynamic User Scenario as described in Section 3.2.2.

4.1. Scenario 1: persistent traffic sources

The data from the first 140 s simulation time is omitted to avoid transient effects. After that statistics are collected for 2000 s simulation time. Simulation results are based on 4–6 replications in each scenario. The 95% confidence interval fraction ratio is below 5%.

4.1.1. Differentiation by a single parameter

This section contains four parts and each parts presents the simulation results of differentiation by one of the four EDCA parameters. The simulations results are presented in two graphs; the left graph presents the results of the scenario where the number of stations is equal for both ACs, the right graph presents the scenario where always exactly 2 AC_0 stations are present and the remaining stations are AC_1 stations.

4.1.1.1. CW_{min}. Parameter CW_{min} determines the size of the interval from which the backoff counter is drawn. A smaller CW_{min} results in smaller backoff counter and a station will reach 0 faster. Because the stations always have packets to transmit, stations with a smaller CW_{min} can transmit more packets.

Fig. 3 presents the results for AC_0 stations with $CW_{min} = 7$ and AC_1 stations with $CW_{min} = 31$. It is indeed seen that stations with a smaller CW_{min} value obtain a larger share of the channel capacity than the other stations. For a small number of stations the ratio between the throughputs of the different ACs is as high as 8, where based on the CW_{min} values a ratio of 4 would be expected. This can be explained as the result of collisions between AC_0 and AC_1 stations. After a collision the involved stations have to double their CW. As AC_0 stations have a smaller CW, they will normally finish their backoff earlier and transmit their packet; this resets the CW and undoes the effects of the collision. AC_1 stations require more time to recover from a collision which negatively influences the ratio. Especially, for a low number of stations this effect is large as most collisions will occur between AC_0 and AC_1 stations, for a higher number of stations collisions will more often occur between AC_0 stations and the ratio drops to the order of 4. The right graph of Fig. 3 confirms this behavior as the ratio remains high for a high number of stations as only two AC_0 stations are present.

An important observation is that the ratio does not grow too large for a high number of stations. This means that AC_1 stations can still obtain channel access and that they will not be starved. Another observation is that the total aggregate saturation throughput decreases for an increasing number of users. Throughput degradation for an increasing number of stations is normal behavior for a WLAN, e.g., see the study of [2]. The comparison of the efficiency reductions of all EDCA parameters is discussed in Section 4.1.3.

4.1.1.2. CW_{max}. The parameter CW_{max} is only effective if a station is involved in a sequence of collisions and CW has grown as large as CW_{max}; CW does not have to be doubled for the following retransmissions.

The left graph of Fig. 4 presents the results for AC_0 stations with $CW_{max} = 31$ and AC_1 stations with $CW_{max} = 1023$. Notice that for AC_0 stations CW_{max} is equal to CW_{min} which means that CW_{max} becomes effective directly after the first collision. For a low number of stations the throughput is almost equal for both ACs as the number of collisions is small. However for higher number of stations, the throughput of AC_1 stations decreases fast and the aggregate throughput of AC_0 even slightly increases. The throughput ratio that starts at 1 also increases fast. This means that AC_0 continuously acquires a larger part of the total capacity and AC_1 is starved. The aggregate throughput of AC_0 starts to decrease for more than 15 stations. It is expected that if the number of AC_0 stations grows larger than the numbers in the graph, then the capacity will drop to 0 fast as AC_0 does not perform any form of exponential backoff. The right graph shows similar behavior; for an increasing number of stations AC_0 stations obtain a larger throughput than

Table 2
System model parameter settings, based on the DSSS PHY layer

WLAN physical		*WLAN MAC*	
Data rate	11 Mbits/s	MAC overhead	224 bits
Basic rate	2 Mbits/s	r_{max}	3
SIFS	10 μs	IFQ length	50 packets
DIFS	50 μs	max # STAS	50
EIFS	304 μs	*TCP/UDP*	
PHY header	48 μs	TCP header	20 Bytes
PLCP header	144 μs	TCP receiver W_{max}	20 packets
Time slot	20 μs	UDP header	20 Bytes
Fixed network		*IP*	
Delay	10 ms	IP header	20 Bytes
Capacity	100 Mbit/s		

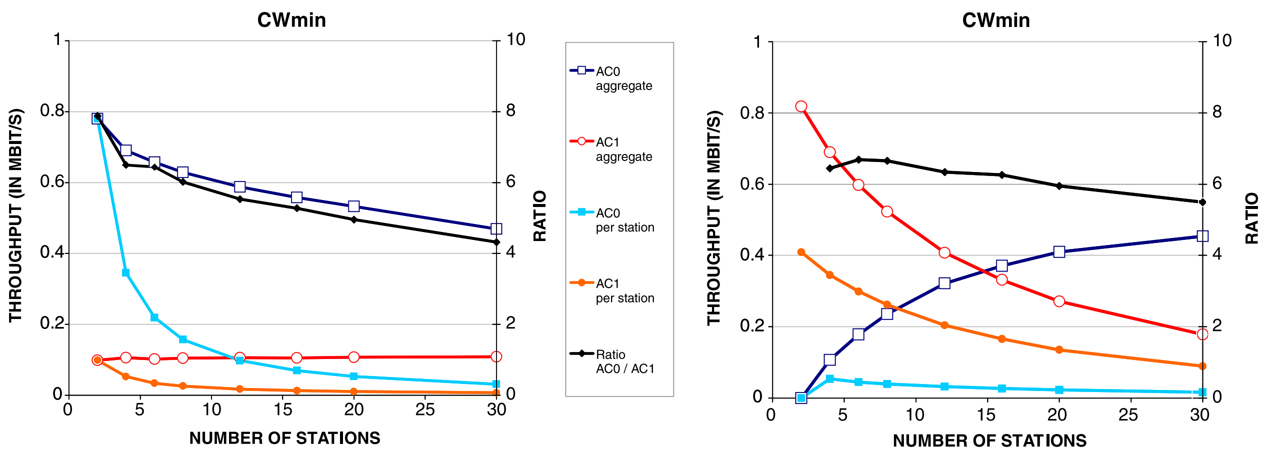

Fig. 3. AC_0 with $\text{CW}_{\text{min}} = 7$ and AC_1 with $\text{CW}_{\text{min}} = 31$. Left, number of stations of both classes is equal. Right, 2 AC_0 stations and the remaining stations are AC_1 stations.

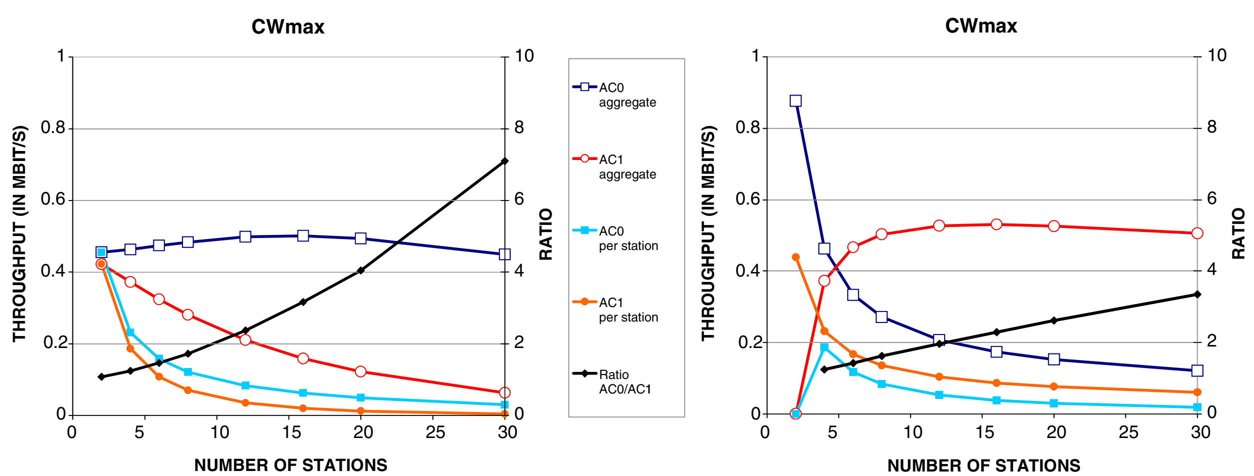

Fig. 4. AC_0 with $\text{CW}_{\text{max}} = 31$ and AC_1 with $\text{CW}_{\text{max}} = 1023$. Left, number of stations equal for both ACs. Right, 2 AC_0 stations, remaining stations are AC_1 stations.

AC_1 stations. However, the effects are less pointed out as there are only 2 AC_0 stations present.

Summarizing, we can say that the differentiating capabilities of CW_{max} appear when many collisions occur. CW_{max} does not provide constant ratio of capacity sharing, e.g., such as CW_{min} does, but it is able to starve out ACs with larger values of CW_{max}. The differentiating capabilities are also at the cost of medium efficiency.

4.1.1.3. $\text{TXOP}_{\text{limit}}$. Note that the $\text{TXOP}_{\text{limit}}$ does not alter the contention behavior of an AC, the $\text{TXOP}_{\text{limit}}$ only determines how long the AC may transmit after it has won a contention.

Fig. 5 presents the results for AC_0 stations with $\text{TXOP}_{\text{limit}} = 0.1$ s. and AC_1 stations with $\text{TXOP}_{\text{limit}} = 0.03$ s. It is easily seen that AC_0 obtains a larger throughput than AC_1. In our simulations, the data payload of each packet is 1460 Bytes and with a transmission rate of 1 Mbit/s this results in the transmission of 2 packets within a $\text{TXOP}_{\text{limit}}$ of 0.03 s. and 7 packets within a $\text{TXOP}_{\text{limit}}$ of 0.1 s. The resulting ratio of the throughputs of stations from AC_0 and AC_1 has the expected value of 3.5 independently of the number of stations. The right graph of Fig. 5 shows the same ratio.

The $\text{TXOP}_{\text{limit}}$ provides a fair differentiating capability that is independent of the number of stations of either class. It is also medium efficient because multiple packets can be transmitted after a single contention.

4.1.1.4. AIFS. The influence of the AIFS is experienced each time the medium becomes free after a transmission and an AC wants to resume its backoff procedure. The other parameters only have an influence on the moments that a new backoff counter has to be drawn or the moment that a contention is won, so we expect that the AIFS is more effective if more stations are present.

Fig. 6 presents the results for AC_0 stations with $\text{AIFSN} = 2$ and AC_1 stations with $\text{AIFSN} = 5$. In the left graph it is seen that even for a small number of users AC_0 stations obtain a larger share than AC_1 stations.

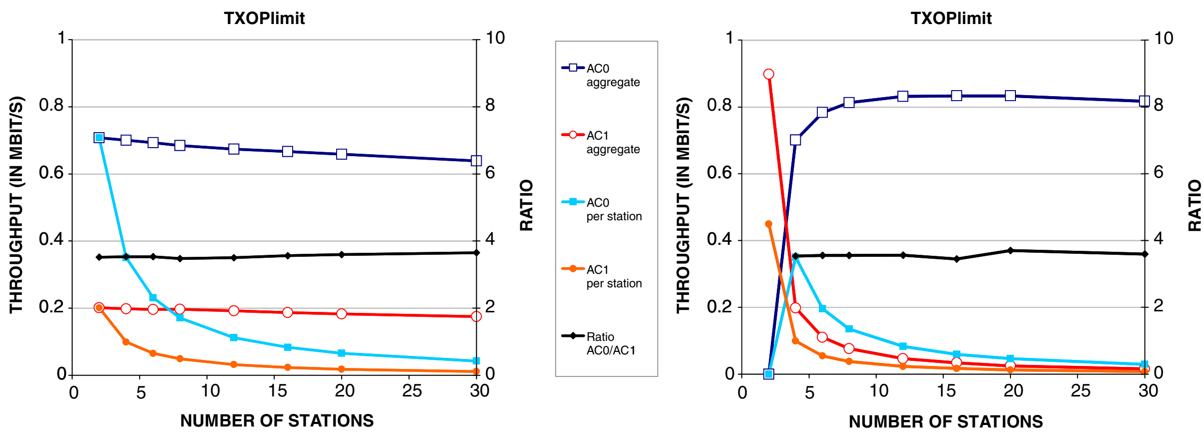

Fig. 5. AC_0 with $TXOP_{limit} = 0.1$ s and AC_1 with $TXOP_{limit} = 0$. Left, number of stations of both classes is equal. Right, 2 AC_0 stations and the remaining stations are AC_1 stations.

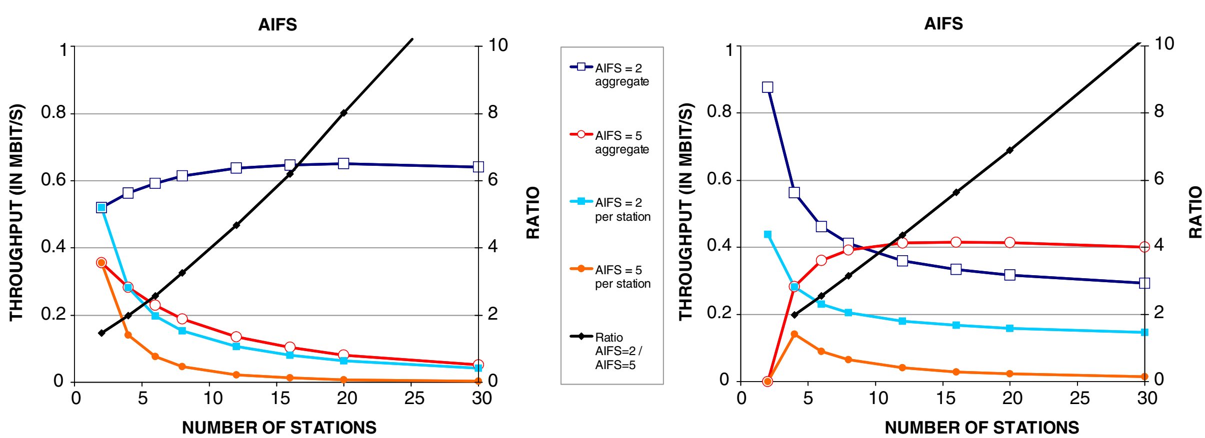

Fig. 6. AC_0 with AIFSN = 2 and AC_1 with AIFSN = 5. Left, number of stations of both classes is equal. Right, 2 AC_0 stations and the remaining stations are AC_1 stations.

If the number of stations increases, this effect becomes more significant. Note that the AIFS of the two ACs differ 3 time slots. This means that each time that a transmission is finished and AC_0 stations can resume their backoff process, AC_1 stations have to wait an additional 3 time slots before they can resume their backoff process. If an AC_0 station initiates a transmission within these first 3 time slots, AC_1 stations did not have the opportunity to decrement their backoff counter. This effect results in the starvation of AC_1 stations for an increasing number of stations. This can be seen by the continuously increasing throughput ratio. The right graph shows a similar behavior although only 2 AC_0 stations are present. During the first 3 time slots after the medium becomes free, the AC_0 stations are only contending amongst themselves which reduces the probability of a collision and increases their throughput.

The parameter AIFS is capable of a stringent differentiation between different ACs. The differentiating capabilities become larger if the number of stations increases and will starve the ACs with larger AIFS.

4.1.2. Differentiation by different parameters

In the previous section the differentiation capabilities of the EDCA parameters were studied in scenarios where only the value of a single parameter was varied. In this section, we also consider two ACs and all parameters are set according to the 802.11B standard, but now for each group a different parameter is set to a more preferable value. Remark that in the previous section AC_0 always had a better performance than AC_1, in the present case it is not a-priori clear which AC has better performance.

Since the $TXOP_{limit}$ does not affect the contention mechanism and consequently the collision probabilities, the parameter is not considered here. The throughput ratio results for the $TXOP_{limit}$ can be obtained directly by multiplying the ratios in the following figures by the ratio of the $TXOP_{limit}$ values of the ACs.

The left graph of Fig. 7 presents the results for AC_0 with $CW_{min} = 15$ and AC_1 with $CW_{max} = 127$. For a small number of stations, AC_0 stations obtain a larger share of the capacity than AC_1 stations. The differentiation of CW_{min} dominates as the number of collisions is small. For an

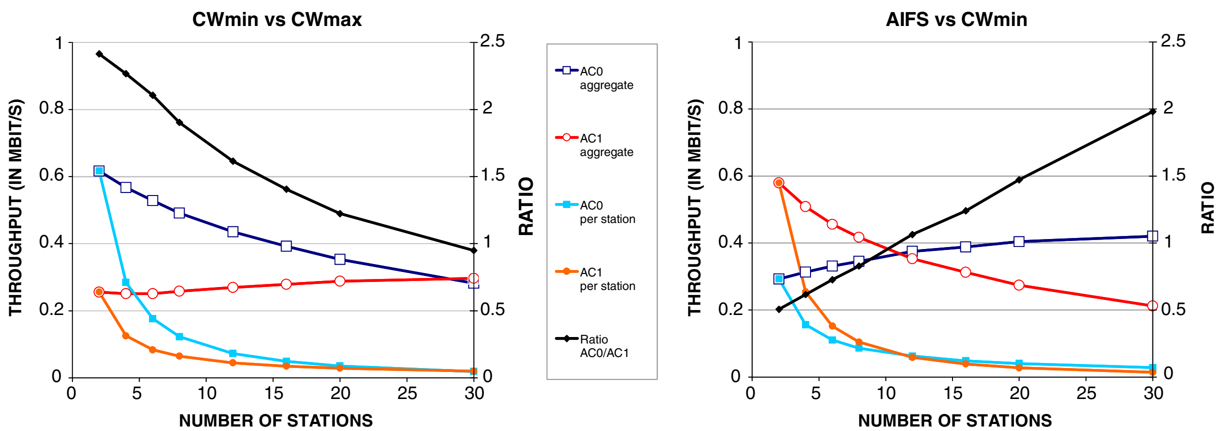

Fig. 7. Left, AC_0 with $CW_{min} = 15$ and AC_1 with $CW_{max} = 127$. Right, AC_0 with $AIFSN = 1$ and AC_1 with $CW_{min} = 15$.

increasing number of stations the ratio between the two ACs decreases as the role of CW_{max} becomes more dominant and around 28 stations the throughput ratio drops below 1, meaning that AC_1 stations obtain a larger share of the capacity than AC_0 stations.

The right graph of Fig. 7 presents the results for AC_0 with $AIFSN = 1$ and AC_1 with $CW_{min} = 15$. On average the backoff window of AC_1 is twice as small as the backoff window of AC_0, but every time that the medium becomes free after a transmission AC_1 has to wait for an extra time slot. For a small number of stations, AC_1 stations obtain a larger throughput (ratio < 1), the number of interrupting transmissions during the countdown of a CW are not too many. However, for an increasing number of stations the differentiating capabilities of AIFS become more dominant and when 11 stations or more are present, AC_0 stations obtain a larger share of the capacity (ratio >1). This behavior was already expected from Figs. 3 and 6 where the ratio of CW_{min} was already slowly decreasing for an increasing number of stations, while the ratio of AIFS was continuously increasing.

Fig. 8 presents the results for AC_0 with $AIFSN = 1$ and AC_1 with $CW_{max} = 127$. In the previous section we have seen that the differentiating capabilities of both CW_{max} and AIFS increase with the number of stations. For a small number of stations the throughput and ratio in Fig. 8 are similar to the results in Fig. 6, although the AIFS values of the ACs differ 3 time slots in Fig. 6 and 1 time slot in Fig. 8. For an increasing number of stations the throughput ratio slightly increases in favor of AC_1, but not as fast as in Fig. 6 indicating that AIFS is more dominant than CW_{max}.

4.1.3. Efficiency of medium usage

In this section, we consider the efficiency of the medium usage under the 802.11E scenarios studied in Section 4.1.1. In particular the aggregate throughput of both AC_0 and AC_1 stations is compared to the aggregate throughput when all stations use 802.11B settings. The results are displayed in Fig. 9.

A small CW_{min} value has a positive effect on the medium efficiency for the case that only a few stations are active. This is due to short backoff periods. However, for an increasing number of stations, the number of collisions increases resulting in a lower saturation throughput.

Parameter CW_{max} is only effective if an AC experiences many collisions. In the figure it is seen that the throughput of the scenario with smaller CW_{max} is almost equal to the

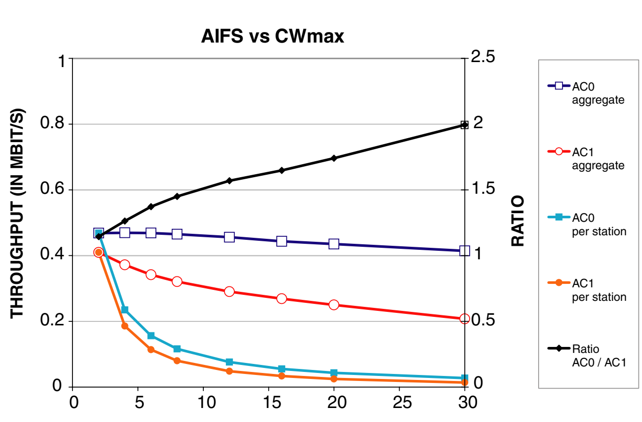

Fig. 8. AC_0 with $AIFSN = 1$ and AC_1 with $CW_{max} = 127$.

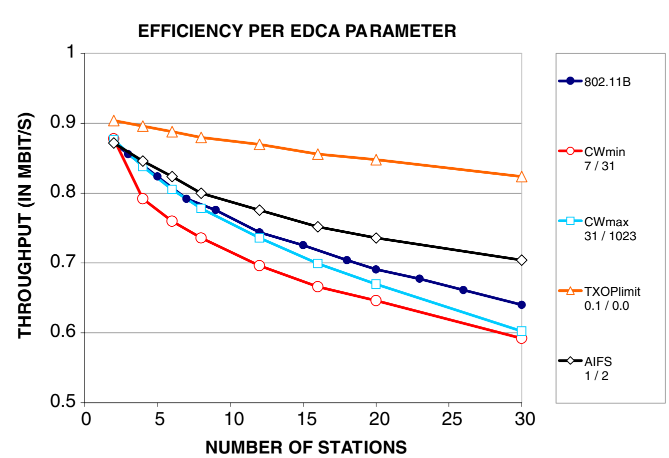

Fig. 9. Efficiency of medium usage. Half of the stations uses 802.11B parameter setting, the other half uses a preferable setting for one EDCA parameter.

throughput of the 802.11B scenario up to 10 stations. For a higher number of stations the reduced CW_{max} hinders an efficient backoff procedure and the aggregate throughput turns out lower.

If $TXOP_{limit}$ is applied, we observe an overall improvement in the medium efficiency, regardless of the number of stations in the network. The $TXOP_{limit}$ does not alter the contention behavior, but when a contention is won, the AC can transmit multiple packets which is more efficient. For transmission of the same amount of data, less contentions are required. The more packets a station is allowed to transmit in one TXOP, the better the throughput performance becomes.

The parameter AIFS improves the aggregate throughput, in particular for a higher number of stations. In this scenario the AIFS values of the groups differ 1 time slot. For the duration of this time slot only AC_0 stations are actually contending and from the curve with all 802.11B stations it can be seen that the system is more efficient with less stations. This improves the overall efficiency and this effect is stronger if the number of stations is larger and also if the difference in AIFS values is larger between the different ACs.

4.1.4. Summary of the results of persistent stations

This section examined the impact of the 802.11E EDCA parameters by means of throughput analysis. The simulations results show the impact of the four EDCA parameters and we summarize these results in Table 3. The parameters are scored using '++' to '−−' on three criteria listed below:

- *Differentiating capability.* Ability to give preference to one Access Category over the other.
- *Fairness.* Ability to share the capacity among the ACs as intended, according to an a-priori defined ratio independent of the number of active users.
- *Efficiency.* Ability to achieve a high aggregate throughput.

The results of our study indicate how the QOS parameters can be applied to meet certain QOS requirements. The $TXOP_{limit}$ has perfect capabilities (cf. Table 3) w.r.t. to all above-mentioned criteria. However, a drawback of setting a large value of $TXOP_{limit}$ is the increase in delay and delay jitter. CW_{min} differentiates well, but for high loads the system capacity decreases. AIFS and CW_{max} both differentiate very well and become even more effective in situations with high load.

4.2. Scenario 2: dynamic user scenario (non-persistent traffic)

This section presents the results of the flow-level simulations described in Section 3.2.2 and specified in Table 1. The traffic settings of Table 1 result in an average of 3 VOIP, 2 VOD and 12 WB users, which corresponds to a load of 0.17. The load is varied by varying the arrival rate of only VOIP or WB. Note that although a 'net' load of 0.17 seems to be light traffic, in fact it is already heavy traffic and the 'gross' load is close to 1. VOIP users have a high gross load caused by inefficient channel usage due to their small packet size. A high number of users also results in a decrease of the channel capacity, so the gross load per user increases.

In this section, we first evaluate the performance of the services for a network with only 802.11B stations. Second we will consider network with 802.11E stations for two different settings of the EDCA parameters. For all networks the load is increased by the load of VOIP, the service class with the highest priority, and for the final EDCA parameter settings we also present the result if the load is increased by WB.

4.2.1. Performance of 802.11B

First, the results of the dynamic scenario over an IEEE 802.11B DCF are presented, thus all service classes have the same priority. The left graph of Fig. 10 shows that the values of the performance metrics of downstream VOIP are worse than for the upstream direction for all loads. Already for low load all three downstream performance metrics are above the QOS targets. The downstream direction performs worse as the majority of all traffic is sent downstream via the AP to the stations. The AP becomes the bottleneck, queueing occurs at its IFQ resulting in larger delays and possibly into packet losses.

The right graph of Fig. 10 presents the transfer times of web Browsers for the same scenario. An increase of the load results in a higher number of VOIP users in the system and WB users' TCP will adapt to the lower remaining available capacity. It is important to realize that VOIP data packets are small and result in low efficiency of the medium usage. So WB users suffer from reduced capacity from both an increasing number of VOIP users and a reducing aggregate throughput.

Fig. 11 presents the results for VOD users. The packet loss is already 0.5% for the lowest load, which is too high

Table 3
Qualitative assessment of the EDCA differentiating parameters

Number of stations	Differentiation		Fairness		Efficiency	
	Low	High	Low	High	Low	High
CW_{min}	++	+	−	++	0	−
CW_{max}	0	++	+	−−	0	−−
AIFS	+	++	0	−	0	++
$TXOP_{limit}$	+	+	++	++	+	++

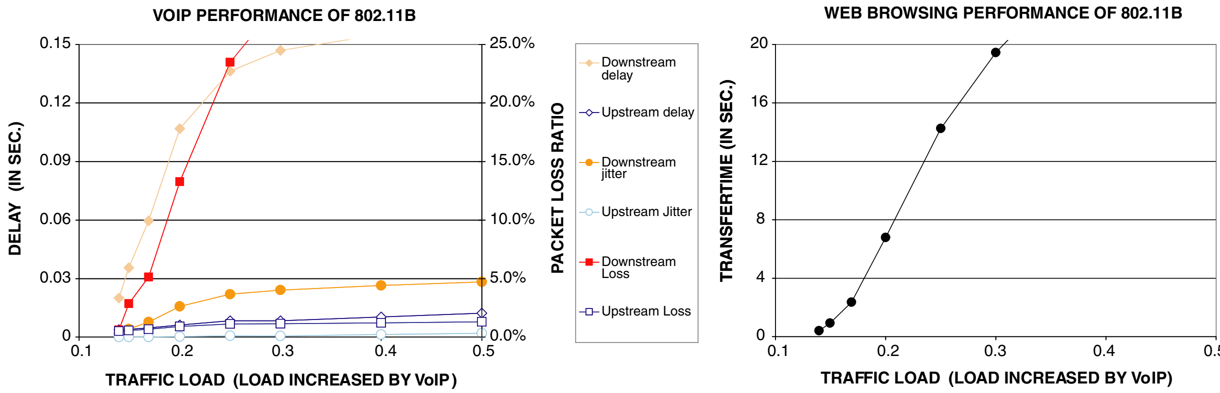

Fig. 10. VOIP and WB performance of 802.11B, load increased by VOIP.

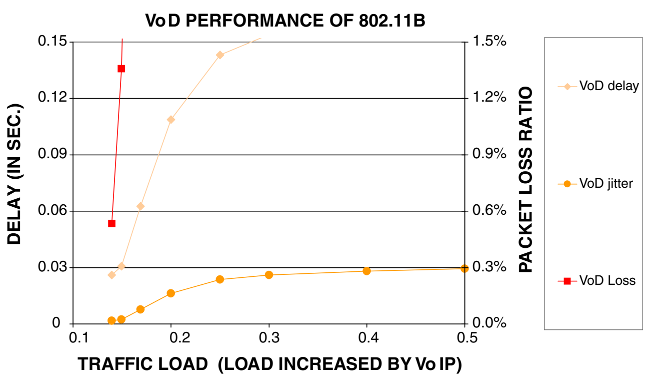

Fig. 11. VOD performance of 802.11B, load increased by VOIP.

Table 4
Parameter settings for EDCA ACS without $TXOP_{limit}$

Traffic	AC_1	AC_2	AC_3
	VOIP	VOD	WB
CW_{min}	7	15	31
CW_{max}	63	255	1023
AIFSN	2	3	4
$TXOP_{limit}$	0	0	0

for video-streaming. The delay and jitter are also quite high, but in practice this can be overcome with the use jitter-buffers; so packet loss should be improved in order to provide VOD with the desired QOS.

4.2.2. Performance of 802.11E (without use of $TXOP_{limit}$)

IEEE 802.11E EDCA differentiates the service classes by mapping them into different Access categories. VOIP, VOD and WB are mapped into the highest, second highest and lowest priority class, respectively.

The values of the differentiation parameters per AC, whose differentiating capabilities are mainly determined by relative differences in their values (e.g. see Section 4.1), are set according to Table 4. To realize the desired QOS for VOIP, all parameters are set to preferable values. Low CW_{min} provides fast contention and fast retransmissions after a collision in order to fulfill the delay constraints, low AIFS also provide fast contention and low CW_{max} provide fast retransmissions. The EDCA parameters for WB are chosen equally to 802.11B (except for AIFSN which is 2 for 802.11B) and the VOD parameters are set in between those of the other Access categories.

Fig. 12 illustrates that the downstream packet loss has improved tremendously compared to situation of all 802.11B stations (note that the scale of the packet loss axis has changed) and that for low loads all performance metrics are within the requirements. For higher loads ($\rho > 0.27$) the downstream packet loss is above 1%, so still

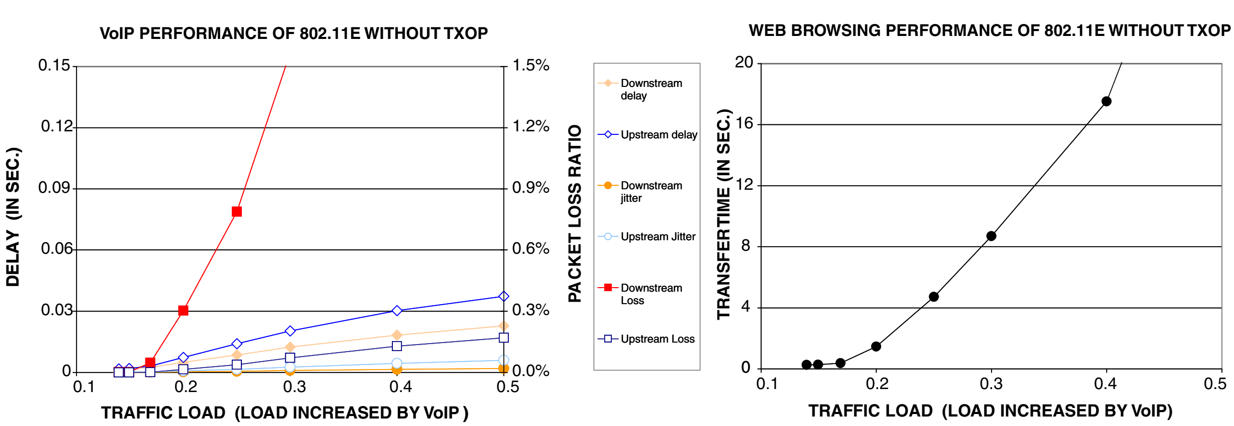

Fig. 12. VOIP and WB performance of 802.11E, load increased by VOIP.

the performance of the downstream direction has to be improved.

The right graph shows that WB traffic, although is has the lowest priority, performs slightly better compared to the DCF. In this scenario the AP has three ACs that are each contending for the medium whereas in the DCF the AP only contends with one AC. This, together with parameter settings, improves the efficiency of the medium usage resulting in better WB performance.

Fig. 13 shows that the results for V OD also improved compared for the situation with 802.11B stations, obviously for the same reasons. For the lowest load (0.15) the packet loss, delay and jitter are all within tolerable levels. However, if the load by VOIP is slightly increased, the performance metrics immediately increase above the QOS targets.

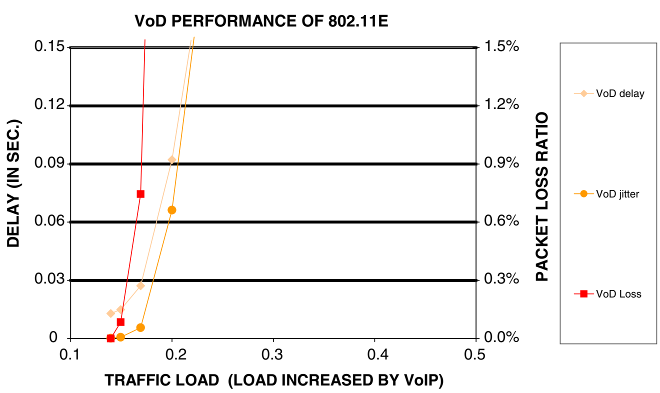

Fig. 13. VOD performance of 802.11E, load increased by VOIP.

Table 5
Parameter s ettings for EDCA ACS with $\text{TXOP}_{\text{limit}}$ for the AP

Traffic	AC_0	AC_1	AC_2	AC_3
	VOIP down	VOIP up	VOD	WB
CW_{min}	7	7	15	31
CW_{max}	63	63	255	1023
AIFSN	2	2	3	4
$\text{TXOP}_{\text{limit}}$	0.06 s	0.03 s	0	0

4.2.3. Performance of 802.11E (with use of $\text{TXOP}_{\text{limit}}$ for the AP)

The previous experiment showed that downstream direction from the AP to the stations is the bottleneck (for all services). To resolve this bottleneck, the capacity for VOIP traffic is increased by setting the $\text{TXOP}_{\text{limit}}$ as specified in Table 5. The AC for VOIP at the AP (AC_0) is set more preferable than the AC for VOIP at the stations (AC_1) since the AP has to serve all stations and requires a larger share of the medium.

The left graph of Fig. 14 shows that downstream packet loss remains the bottleneck and the performance is only slightly improved. The improvement results from the AP that can transmit multiple VOIP packets per won contention; although it wins the same number of contentions, the AP can empty its buffer faster. The downstream improvements are at the cost of extra delay and delay jitter for VOIP in the upstream direction, but they are still within the requirements. The performance of lower priority services WB and VOD have also decreased slightly, cf. the right graph of Figs. 14 and 15.

In the previous three figures the load was always increased by increasing the load of the highest priority AC. Fig. 16 presents the results for the same scenario, but

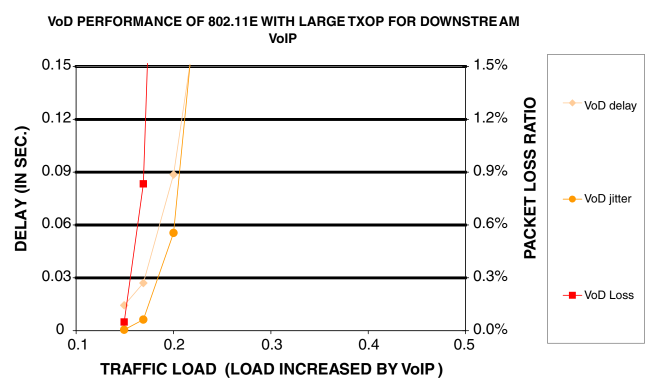

Fig. 15. VOD performance of 802.11E with $\text{TXOP}_{\text{limit}}$ for downstream VOIP, load increased by VOIP.

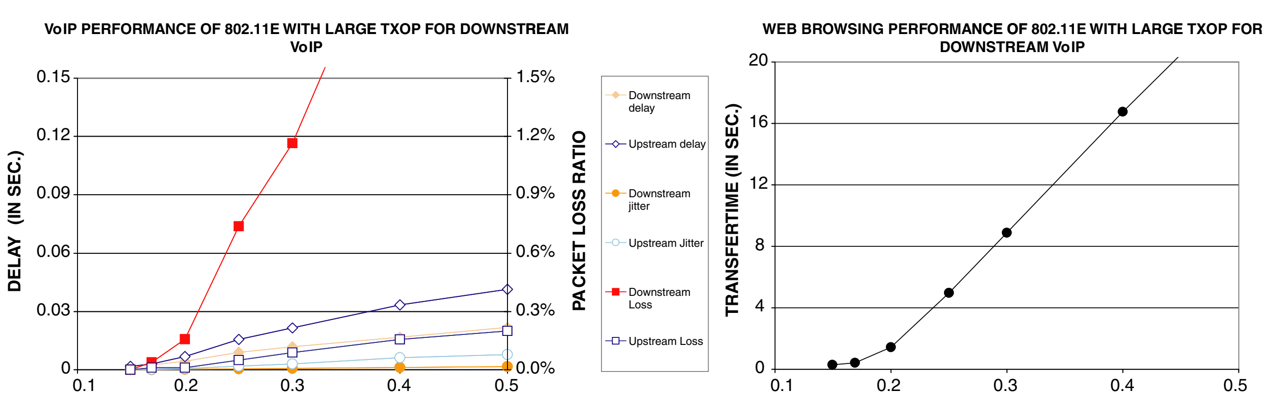

Fig. 14. VOIP and WB performance of 802.11E with $\text{TXOP}_{\text{limit}}$ for downstream VOIP, load increased by VOIP.

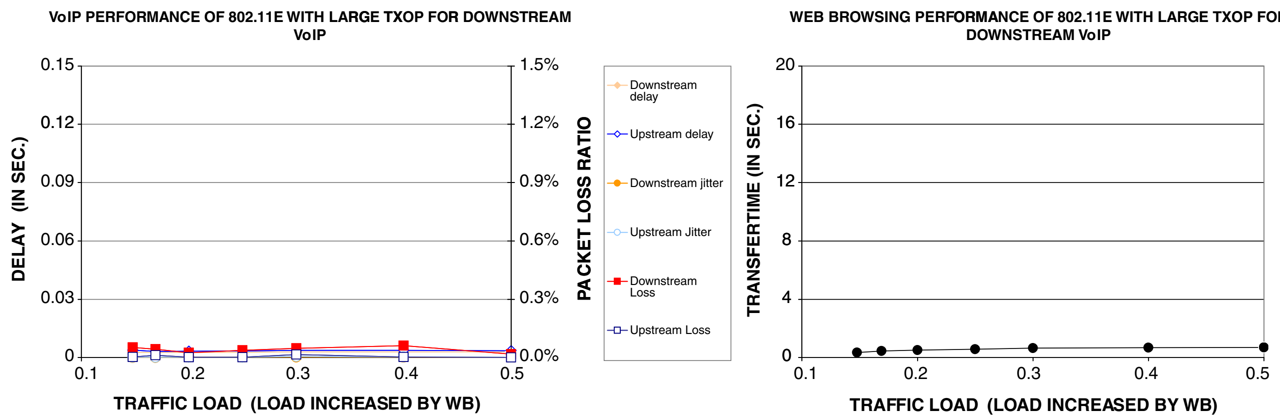

Fig. 16. VoIP and WB performance of 802.11E with $\text{TXOP}_{\text{limit}}$ for downstream VoIP, load increased by WB.

now the load is increased by WB traffic. Fig. 16 shows that the performance of VoIP, as the highest priority AC, is hardly influenced if the load is increased by WB traffic. The right graph shows that the transfer time of WB almost does not increase. If the load is increased above a value of 0.3, transfer times even slightly decrease. The reason is that the load becomes too high and the system reaches it maximum number of users due to the CAC. If the load is increased further, on average more WB and less VoIP will enter the system which changes the traffic mixture. As a service VoIP puts a large strain on the system capacity due to the small packets, so with the changing mixture of services present for increasing load, the aggregate system throughput increases resulting in shorter transfer times. For VoD the same behavior is seen even more clear as delay, jitter and packet loss decrease for increasing load, cf. Fig. 17.

The results in this section illustrate that tuning of the EDCA parameters can improve the performance of a desired AC in realistic traffic scenarios. However, improving the performance of a desired AC will be at the cost of the performance of the other ACs, especially lower priority ACs. Thus, the 'optimal' choice of the QoS differentiation parameters settings depends on the specific objectives of the operator; e.g., VoIP protection (because of its high delay sensitivity) or guaranteed throughput for web browsers.

5. Concluding remarks

In this paper, we have studied the EDCA mechanism for QoS provisioning in WLAN. First, extensive simulations of scenarios with persistent users illustrate the QoS differentiating capabilities of the EDCA parameters. The impact of the EDCA parameters is not ambiguous and depends on the system characteristics, e.g., the number and types of users. The results are summarized in Table 3.

Second, we have studied the impact of the EDCA parameters in a scenario with three service types and dynamic arrivals and departures of users. It is shown that the plain 802.11B is not capable of fulfilling service requirements of interactive services. 802.11E improves the performance, however a major drawback is that the Access Point becomes

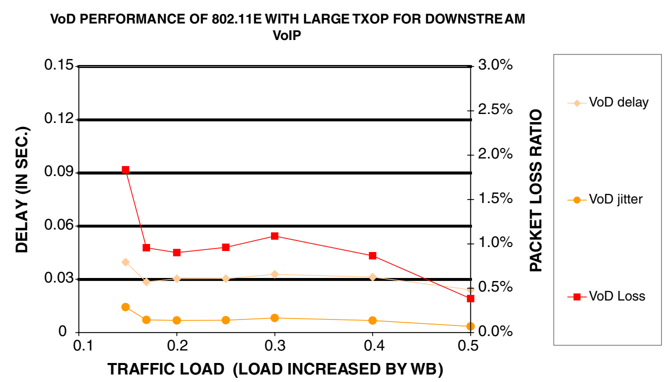

Fig. 17. VoD performance of 802.11E with $\text{TXOP}_{\text{limit}}$ for downstream VoIP, load increased by WB.

the bottleneck in the downstream direction. The performance can be improved by setting the EDCA parameter values of the AP to a more preferable value than the value of the corresponding ACs of the stations.

The EDCA is only capable of providing service differentiation, and not of delivering absolute QoS guarantees. The best approach to attempt providing absolute guarantees is to deploy Call Admission Control (CAC). The results of the present study can be used to determine CAC boundaries on the number of users per type that may be present in the system.

Acknowledgments

This work has been partially funded by the Dutch Ministry of Economics Affairs under the programs "Technologische Samenwerking ICT-doorbraakprojecten", project TSIT1025 BEYOND 3G and "Innovatiesubsidie Samenwerkingsprojecten", Project IS043014 EASY WIRELESS.

References

[1] I. Aad, C. Castelluccia, Priorities in WLANS, Comput. Netw. vol. 41 (no. 4) (2003) 505–526.

[2] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, IEEE J. Selec. Areas in Commun. vol. 18 (no. 3) (2000) 535–547.
[3] S.K. Cheung, J.L. van den Berg, R.J. Boucherie, R. Litjens, F. Roijers, An analytical packet/flow-level modelling approach for wireless LAN with Quality-of-Service support, Proceedings of ITC 19, Beijing, China, 2005.
[4] S. Choi, J. del Prado, S. Shankar N, S. Mangold, IEEE 802.11E Contention-Based Channel Access (EDCF), Proceedings of ICC 2003, Anchorage AL, USA, 2003.
[5] C.H. Foh, M. Zukerman, Performance analysis of the IEEE 802.11 MAC protocol, Proceedings of European Wireless '02, Florence, Italy, 2002.
[6] P. Garg, R. Doshi, R. Greene, M. Baker, M. Malek, X. Cheng, Using IEEE 802.11E MAC for QOS over Wireless, Proceedings of the IPCCC 2003, Phoenix, Arizona, 2003.
[7] D. He, Ch. Shen, Simulation Study of IEEE 802.11E EDCF, IST Project Moby Dick.
[8] IEEE P802.11B/D7.0, Supplement: higher speed physical layer extension in the 2.4 GHz band, 1999.
[9] IEEE P802.11E/D11.0, Draft supplement: enhancements for Quality of Service, October 2004.
[10] J. Kim, C. Kim, Performance analysis and evaluation of IEEE 802.11E EDCF, Wireless Commun. Mobile Comput. 4 (2004) 55–74.
[11] A. Lindgren, A. Almquist, O. Schelen, Evaluation of Quality of Service Schemes for IEEE 802.11 Wireless LAN, Proceedings of IEEE LCN, 2001, pp. 348–351.
[12] R. Litjens, F. Roijers, J.L. van den Berg, R.J. Boucherie, M.J. Fleuren, Performance analysis of wireless LANS: an integrated packet/flow level approach, Proceedings of ITC 18, Berlin, Germany, 2003.
[13] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, L. Stibor, IEEE 802.11E Wireless LAN for Quality of Service, Proceedings of European Wireless 2002, Florence, Italy, 2002.
[14] Network Simulator 2: <http://www.isi.edu/nsnam/ns/>.
[15] <http://www.tkn.tu-berlin.de/research/802.11e_ns2/>.
[16] T. Raimondi, M. Davis, Design rules for a class-based differentiatied service QOS scheme in IEEE 802.11E wireless LANS, Proceedings of MSWiM'04, Venice, Italy, 2004.
[17] J. Zhao, Z. Guo, Q. Zhang, W. Zhu, Performance study of MAC for service differentiation in IEEE 802.11, Proceedings of IEEE GLOBECOM 2002, Taipei, Taiwan, 2002.

Frank Roijers received his M.Sc. degree in Business Mathematics and Informatics from the Vrije Universiteit Amsterdam (VU) in 1998. In 2000 he joined TNO Information and Communication Technology (TNO ICT), the former KPN Research. His research interests include performance analysis and traffic engineering of communication networks (IP, Wireless LAN, ad-hoc networks) and queueing theory. Since 2004 he is working towards a Ph.D. in a joint project of TNO ICT and the Centre for Mathematics and Computer Science (CWI), Amsterdam.

Hans van den Berg received the M.Sc. and Ph.D. degree in mathematics (stochastic operations research) from the University of Utrecht, the Netherlands, in 1986 and 1990 respectively. From 1986, he worked at the Centre for Mathematics and Computer Science (CWI), Amsterdam. In 1990, he joined KPN Research (now TNO Information and Communication Technology, since January 2003). He is particularly working on performance analysis, traffic management and QoS provisioning for wired and wireless multi-service communication networks (ATM, IP, UMTS, WLAN, ad-hoc networks). Currently, he is a Senior Research Member and leader of the QoS group within TNO ICT, Delft, The Netherlands. Since July 2003 Hans van den Berg has a part-time position as full professor within the faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente.

Xiang Fan was a master student at the department of Design and Analysis of Communications Systems (DACS) at the University of Twente. In 2004 he concluded an 8 months internship at TNO ICT with his thesis "QoS control in Wireless Local Area Networks" and he received his M.Sc. in Electrical Engineering. Currently he is employed at Delta Electronics where he is working as a Field Application Engineer for the product group 'Fans'.

Maria Fleuren received her M.Sc. degree in Applied Physics from Delft University of Technology in 1995 and her M.Sc. degree in Mathematics from Leiden University in 1997. In 1998 she joined KPN Research which in 2003 became TNO Information and Communication Technology (TNO ICT). Her research focus is on performance analysis of mobile communication networks, especially radio resource management techniques for UMTS, HSDPA and WLAN.