Understanding 802.11e contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations

Giuseppe Bianchi(+), Ilenia Tinnirello(*), Luca Scalia(*)

(+) Università degli Studi di Roma - Tor Vergata, Dipartimento di Ingegneria Elettronica
Via del Politecnico, 1, 00133 Roma, Italy
giuseppe.bianchi@uniroma2.it

(*) Università degli Studi di Palermo, Dipartimento di Ingegneria Elettrica
Viale delle Scienze, 90128, Palermo, Italy
ilenia.tinnirello@tti.unipa.it, luca.scalia@tti.unipa.it

***Abstract*— The IEEE 802.11e task group has reached a stable consensus on two basic contention-based priority mechanisms to promote for standardization: usage of different Arbitration Inter Frame Spaces (AIFS), and usage of different minimum/maximum contention windows. Goal of this paper is to provide a thorough understanding of the principles behind their operation. To this purpose, rather than limit our investigation to high-level (e.g. throughput and delay) performance figures, we take a closer look to their detailed operation also in terms of low-level performance metrics (such as probability of accessing specific channel slots). Our investigation on one side confirms that the AIFS differentiation provides superior and more robust operation when compared to contention window differentiation. On the other side, it highlights performance issues related to the coexistence between 802.11e contention-based stations with legacy 802.11 stations, and provides guidelines for the 802.11e parameter settings when such a coexistence is aimed at.**

I. Introduction

The IEEE 802.11 technology [1] is experiencing an impressive market success. Cheap and easy-to-install components, unlicensed spectrum, broadband capabilities, interoperability granted by standards and certifications (e.g. WiFi): these are a few of the key factors that are driving the evolution of WLAN from niche technology to public access mean. The present challenge of WLANs is to offer a large portfolio of wireless mobile services to highly heterogeneous users with widely different requirements. This goal can be accomplished only by introducing suitable forms of service differentiation support at the various level of a complex wireless network architecture.

A basic building block for service differentiation is the introduction of layer-2 prioritized delivery mechanisms for different traffic classes (and users), and support of quality of service objectives. In switched Ethernet networks, service differentiation is managed within switches, through IEEE 802.1p priorities / Virtual LAN IEEE 802.1Q tags. In IEEE 802.11 WLANs, all stations share the access to the same radio channel and no switching operation is possible. Thus, the service differentiation mechanisms must be compulsorily introduced as MAC layer extensions.

For this reason, the task-group 802.11e has been established in July 1999, chartered to introduce Quality of Service (QoS) support at the MAC layer. The current 802.11e draft standard (version 10.0 [2]) defines two mechanisms, Enhanced Distributed Channel Access (EDCA) and Hybrid Coordination Function Controlled Channel Access (HCCA), both backward compatible with the legacy Distributed Coordination Function (DCF) access mechanism, defined by the 1999 standard [1]. A thorough presentation of the main features of 802.11e is provided in companion papers within this special issue [3], [4].

In this paper we focus on the performance effectiveness of the priority mechanisms defined in the EDCA specification. EDCA considers two basic priority mechanisms for accessing the channel: different per-class setting of the contention window backoff parameters (CW_{min} and CW_{max}), and different per-class setting of the idle time after which a transmission may occur (Arbitration Inter-Frame Space - AIFS). Once a station accesses the channel, EDCA also provides the ability to differentiate the time interval a station is authorized

This work has been partially supported by the Italian Ministry of Education, University and Research (MIUR) in the frame of the Italian project PRIN-TWELVE

to hold the channel (Transmission Opportunity - TXOP). Since the thorough understanding of the impact of different TXOP settings on service differentiation is somewhat immediate, and since the TXOP feature is not mandatory and can be disabled, in the following we do not analyze this supplementary mechanism.

The performance evaluation of 802.11e/EDCA has been thoroughly carried out in recent literature. Different aspects have been investigated, including the coexistence of data, voice and video applications [5], [6]; the need to integrate MAC-level service differentiation mechanisms with admission control policies [7], [8]; the system capacity evaluation and the impact of different MAC parameter settings [9], [10], and so on. All these works, as well as many other 802.11e/EDCA performance studies appeared in the literature, here not mentioned for reasons of space, derive performance figures such as (per-class) throughput, delay statistics and fairness indexes. These performance indicators have a fundamental impact in terms of system dimensioning and parameters engineering, since they quantify the performance experienced by the customers of an 802.11e network.

This paper takes a completely different point of view. Rather than providing system engineering insights, our goal is to conduct the reader to thoroughly understand the principles and the physical reasons behind the operation of the service differentiation mechanisms proposed in EDCA (namely, AIFS and CW$_{min}$ differentiation). To this purpose, we mainly focus our study on low-level performance figures which are "internal" to the system operation (e.g. probability that specific slot times are occupied by MAC frames of given traffic classes), rather than limiting our investigation to high-level performance figures. Our proposed low-level view results helpful not only to provide a better understanding of the effectiveness of various EDCA mechanisms, but also to better assess more detailed technical issues, such as the important issue of coexistence between legacy DCF stations and EDCA terminals.

II. CW$_{min}$ AND AIFS DIFFERENTIATION

The EDCA proposal of the IEEE 802.11e Task Group is devised to differentiate the channel access probability among different traffic sources. As explained into greater details in [3], [4], packets arriving to the MAC (Mac Service Data Units, MSDUs) are mapped into four different access categories (ACs), which represent four different levels of service for the contention to the shared medium. Each AC contends to the medium with the same rules of standard DCF, i.e, wait until the channel is idle for a given amount of Inter Frame Space IFS, and then access/retry following exponential backoff rules. The access probability differentiation is provided by using i) different Arbitration Inter-Frame Spaces (AIFS), instead of the constant Distributed InterFrame Space (DIFS) used in DCF, and ii) different values for the minimum/maximum contention windows to be used for the backoff time extraction. Then, each AC is specified by the values AIFS[AC], CW$_{min}$[AC], and CW$_{max}$[AC]. The AIFS[AC] values differ each one for an integer number of backoff slots. In particular, AIFS[AC]=AIFSN[AC]· aSlotTime + aSIFSTime, where AIFSN[AC] is an integer greater than 1 for normal stations and greater than 0 for APs.

Access Category	CW$_{min}$	CW$_{max}$	AIFSN
AC_BK	aCW_{min}	aCW_{max}	7
AC_BE	aCW_{min}	aCW_{max}	3
AC_VI	$aCW_{min}/2$	aCW_{min}	2
AC_VO	$aCW_{min}/4$	$aCW_{min}/2$	2

TABLE I
EDCA DEFAULT SETTINGS

Table I shows the default values of the channel access parameters defined in EDCA for the four ACs (BK=Background, BE=Best Effort, VI=Video, VO=Voice). Note that these parameters are not fixed: in each beacon frame, the AP broadcasts the values chosen for each AC. Indeed, these values may be also dynamically adapted according to the network conditions. Obviously, the smaller the AIFSN[AC] and the CW$_{min}$[AC], the higher is the probability to win the contention with the other ACs. Separate queues are maintained in each station for different ACs and each one behaves as a single enhanced DCF contending entity. When more than one AC of the same station expires its backoff counter, a virtual collision occurs and the highest priority packet among the colliding ones is selected for actual transmission on the radio channel.

In the following, after discussing the fairness property of standard DCF, we introduce and separately analyze into details both CW$_{min}$ and AIFS differentiation. These two prioritization mechanisms are individually evaluated considering as a reference the legacy DCF access, and specifically comparing the performance obtained when EDCA stations compete with standard DCF ones. This allows not only to understand the effect of the service differentiation parameters, but also to tackle the somewhat tricky issue of coexistence with legacy DCF stations. For sake of simplicity, we will assume that each EDCA station supports a single access category, and thus no virtual collision may occur inside the EDCA station.

A. *Fairness issues in standard DCF*

According to the IEEE 802.11 Distributed Coordination Function (DCF) rules [1], a station with a new frame

(MPDU) to transmit monitors the channel activity. If the channel is sensed idle for a period of time equal to a Distributed Inter Frame Space (DIFS), the station transmits. Otherwise, if the channel is sensed busy (either immediately or during the DIFS), the station persists to monitor the channel until it is measured idle for a DIFS. At this point, the station generates a random backoff interval before transmitting, to minimize the probability of collision with packets being transmitted by other stations. In addition, to avoid channel capture, a station must wait a random backoff time between two consecutive packet transmissions, even if the medium is sensed idle in the DIFS time.

DCF adopts an exponential backoff scheme, and employs a discrete-time backoff scale. The time immediately following an idle DIFS is slotted, and a station is allowed to transmit only at the beginning of each *Slot Time*. At each MPDU transmission, the backoff time is uniformly chosen in the range $(0, W-1)$. At the first transmission attempt for a given MPDU, the value $W-1$ is set equal to a parameter $CW_{\rm min}$, called Minimum Contention Window. CW$_{min}$ is equal for all stations, and set to the value 31 in the case of 802.11b. After each unsuccessful transmission, W is doubled, until $W-1$ reaches a maximum value CW$_{\rm max}$, called Maximum Contention Window, and equal to 1023 in the 802.11b case.

Assuming that all stations receive the same channel quality (or, more restrictively, that ideal channel conditions occur), this operation has been shown to be long-term fair [11] in terms of access probability. This means that, in average, the same number of *successful* channel accesses is granted to each contenting station or, equivalently, that over a long time interval, stations which have always a frame ready for transmission, will deliver the same amount of MPDUs).

It is interesting to note that long-term fairness can be derived as a corollary of the fact that, in the absence of transmission errors, all stations experience the same probability of collision [12]. In order to understand this statement, we recall from [12] that, in the assumption of greedy (saturated) traffic sources, the DCF channel accesses can be considered as composed of slot times of uneven size. A channel slot may be either empty, and thus lasts a slot-time as specified by the standard (20 μs in the case of 802.11b), or busy, and thus lasts the amount of time necessary to complete a frame transmission (or a collision among two or more frames), plus the extra DIFS time necessary to permit other stations to access the channel again. Clearly, the time elapsing between two consecutive successful transmissions is related to the backoff parameters used by the station (namely, CW$_{min}$ and CW$_{max}$) and to the probability that a transmitted frame collides[1]. Thus, as long as all stations encounter the same collision probability, and employ the same backoff parameters, their time-averaged throughput performance is the same.

B. *CW$_{min}$ differentiation*

The idea behind the CW$_{min}$ differentiation employed in EDCA is to change the amount of transmission opportunities provided to each traffic class. A station with a lower value of the contention window will, in average, reduce the time needed to successfully deliver a packet, and thus will experience improved performance in comparison to stations with higher contention window values. The average value of the contention window can be tuned through the differentiated setting of the backoff parameters, and specifically of the minimum and maximum contention window values CW$_{min}$ and CW$_{max}$[2]. In practice, in low network congestion situations, changes in the CW$_{max}$ parameter have limited effects on throughput differentiation. For example, if we assume that the probability of collision is negligible, a station, regardless of its CW$_{max}$ value, will successfully transmit, in average once every CW$_{min}/2+1$ slots, which correspond to the average number CW$_{min}/2$ of backoff slots plus the slot used for transmission. Clearly, a station employing a double CW$_{min}$ value, will receive (if the collision probability is small), about one half of the transmission opportunities of the other stations.

Figure 1 shows throughput results in a scenario in which N legacy 802.11b DCF stations share the channel with the same number N of EDCA stations. In order to be granted priority over the DCF stations, EDCA must be configured with CW$_{min}$ values smaller than

[1]More formally, if p is the probability of collision, and neglecting the effect of the retransmission limit, the average time (measured in slots as defined above) between two consecutive successful transmissions is readily obtained as

$$\frac{W_0}{2}+1+p\left(\frac{W_1}{2}+1+p\left(\frac{W_2}{2}+1+\cdots\right)\right)$$

where $W_i = \min\left(2^i(CW_{min}+1)-1, CW_{max}\right)$ is the backoff window used for the i-th retransmission, and $W_i/2+1$ is the average duration, in slots, of the i-th backoff period including the subsequent transmission slot. For p small, this expression can be approximated as $CW_{min}/2+1$.

[2]In early versions of the 802.11e draft specifications, a further parameter considered for differentiation was the Persistence Factor, i.e. the multiplicative factor for the contention window increment after a collision (which is set to 2 in the binary exponential backoff used by DCF). But it was soon understood that its effect is similar - though less effective - to the change of just the minimum and maximum contention window values. Hence it was abandoned in later standardization stages.

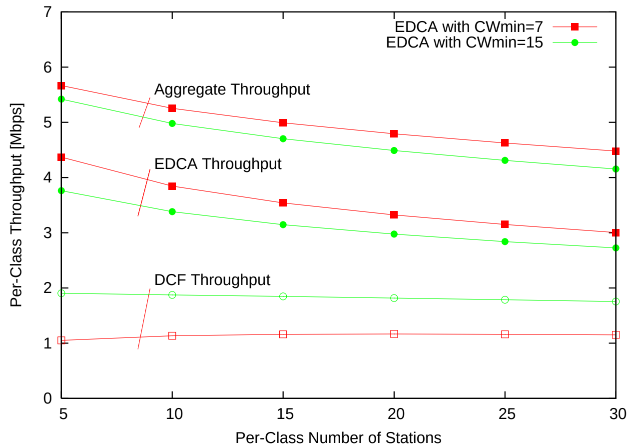

Fig. 1. DCF versus EDCA throughput with CW$_{min}$ differentiation

the legacy DCF value CW$_{min}$ $= 31$. In the figure we have thus chosen CW$_{min}$=7 and 15. EDCA stations have been configured with an AIFSN value equal to 3, which approximates (for reasons that will become clear in the following sections) the DIFS setting for legacy DCF stations. In both DCF and EDCA cases, CW$_{max}$ $= 1023$. The packet size has been fixed to 1500 bytes (Ethernet MTU) and the retransmission limit is set to 7 for all the stations. Control frames are transmitted at a basic rate equal to 2 mbps, while the MPDU is transmitted at 11 mbps. We measured performance in saturation conditions. Although this assumption is not realistic for real-time application, it represents a very interesting case study to derive the limit performance, i.e., the maximum amount of bandwidth that high priority ACs can obtain sharing the channel with best effort DCF stations.

The figure shows that, as expected, for low values of N, the share of resources between EDCA and DCF stations is inversely proportional to the employed CW$_{min}$ value. For example, in the case of $N = 5$, we see that the throughput performance of EDCA, when CW$_{min}$ $= 7$, is about 4 times the corresponding throughput performance of DCF (which uses CW$_{min}$ $= 31$), and, similarly, when CW$_{min}$ $= 15$, it is about the double than the DCF throughput.

We note that, as the number of competing stations grow, the EDCA throughput significantly reduces, while the DCF one decreases only slightly. This phenomenon is not desirable, since performance degradations due to the network congestion should be attributed first to best effort stations.

Finally, the figure shows that smaller CW$_{min}$ values lead to a smaller aggregate throughput. This is an obvious drawback of CW$_{min}$ differentiation: the performance differentiation is paid in terms of aggregate performance. This phenomenon is easily explained by considering that

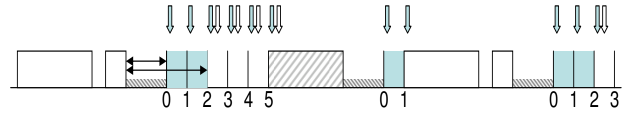

Fig. 2. Slotted channel accesses and protected slots example

the reduction of the minimum contention window value may significantly increase the probability of collision on the channel, thus reducing the overall effectiveness of the random access mechanism.

C. AIFS differentiation

AIFS differentiation is motivated by a completely different (and somewhat more complex) physical rationale. Rather than differentiating the performance by changing the backoff structure (through different settings of the CW$_{min}$ and CW$_{max}$ parameters), the idea is to reserve channel slots for the access of higher priority stations.

This is accomplished by using different Arbitration InterFrame Space (AIFS) values for different traffic classes. The AIFS is the amount of time a station defers the access to the channel following a busy channel period. Once an AIFS has elapsed, the station access is managed by the normal backoff rules. Figure 2 graphically illustrates the AIFS differentiation mechanism for the case of two traffic classes. After every busy channel period, each station waits for a time equal to its AIFS value. If, as in the case considered in the figure, the AIFS values are different, there exists a period of time in which the stations with shorter AIFS value (namely, the higher priority stations) may access the channel, while the stations with longer AIFS (lower priority stations) are prevented from accessing the channel.

The EDCA specification imposes that AIFS values differ for an integer number of slot times. This implies that the channel access can be still considered as slotted, and stations may access the channel only at the discrete time instants indicated in figure 2 by arrows. We note that there exist some discrete instants of time, hereafter referred to as ”protected slots”, where only high priority stations may access the channel. In the figure, the protected slots are shaded and indicated by a single arrow. A low priority station can access the channel only if no high priority station has accessed the channel in one of the previous protected slots (in the example considered, the difference between the AIFS values has been set to 2, and the protected slots are the ones indexed as 0 and 1, while slots numbered from 2 may be accessed by both classes).

A fundamental issue of the AIFS differentiation is that protected slots occur after every busy channel period.

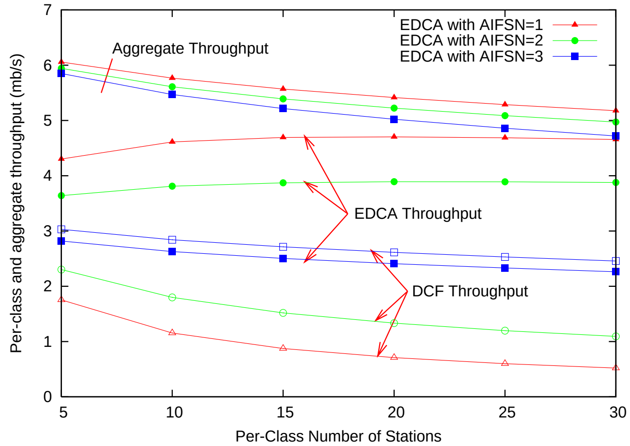

Fig. 3. DCF versus EDCA throughput with AIFS differentiation

This implies that the percentage of protected slots significantly increases as long as the network congestion increases. In fact, a greater number of competing stations implies that the average number of slots between consecutive busy channel periods reduces, and thus the fraction of protected slots over the total number of idle slots gets larger.

Figure 3 shows throughput results in a scenario in which N legacy DCF stations share the channel with the same number N of EDCA stations. EDCA stations have been configured with the standard DCF backoff parameters (CW$_{min}$ = 31 and CW$_{max}$ = 1023). All the other simulation settings are the same of figure 1.

First, figure 3 shows that EDCA stations configured with an AIFSN value equal to 3 achieve performance close to that of legacy DCF stations. This counter-intuitive result (we recall that a DIFS is equal to an AIFS with AIFSN=2) requires a detailed analysis of the backoff counter decrement mechanism used in EDCA, and will be justified in the second part of this paper.

Moreover, figure 3 shows that performance depends dramatically on the AIFSN setting. It is remarkable to note that, by lowering the EDCA AIFSN setting of just one slot, AIFS differentiation shows an impressive effectiveness in protecting the EDCA priority traffic from the legacy DCF one, especially in the presence of network congestion (e.g. see the case of 30+30 stations). Indeed, AIFS differentiation correctly reacts to network congestion by penalizing DCF stations, while EDCA stations do not experience an aggregate throughput reduction (which, on the contrary, actually shows a slight increase[3]).

The analysis of the aggregate AIFS throughput curves also allows to draw a very interesting conclusion: contrarily to the CW$_{min}$ differentiation, *the AIFS mechanism is beneficial in terms of aggregate throughput performance*. This is a direct consequence of the fact that AIFS differentiation introduces protected slots in which a lower number of stations compete, thus increasing the effectiveness of the overall random access mechanism.

III. A CLOSER LOOK

The previous discussion has highlighted the different operation and behaviour of AIFS and CW$_{min}$ differentiation, and has shown how these translate into high-level performance figures (specifically, throughput performance have been used as benchmark). In order to understand how these mechanisms operate in scenarios where EDCA stations compete with legacy DCF terminals, it is necessary to take a closer look to some further technical details and provide additional performance insights in terms of low-level performance metrics.

A. Backoff counter decrement rules

EDCA slightly differs from DCF in terms of how the backoff counter is managed (decremented, frozen, resumed). However, such apparently minor difference (which might perhaps appear as a technicality) indeed has some important consequences in terms of performance of EDCA access categories, especially when they compete with legacy DCF stations.

In standard DCF, the backoff counter is decremented at each idle slot-time, frozen during channel activity periods, and resumed after the medium is sensed idle again for a DIFS interval. This implies that a legacy DCF station, after a DIFS, resumes the backoff counter to the discrete value the station had at the instant of time the busy channel period started. An illustrative example is shown in figure 4. Here, a busy channel period (i.e. a transmission from one or more other stations) starts while the backoff counter of the considered DCF station is equal to 4. This value will be frozen during the busy channel period, and will be resumed, again to the value 4, only a DIFS after the end of the busy period. As a consequence, it will be decremented to the value 3 only a slot after the DIFS.

[3]However, the reader should be careful to note that the aggregate throughput is shared among the EDCA stations, and thus, as the number of EDCA stations grows, the per-station throughput ultimately reduces. If a minimum rate must be guaranteed to each station (such as in the case of voice or video flows), solutions devised to enforce an upper-bound on the number of competing stations (e.g. Admission Control mechanisms) are necessary.

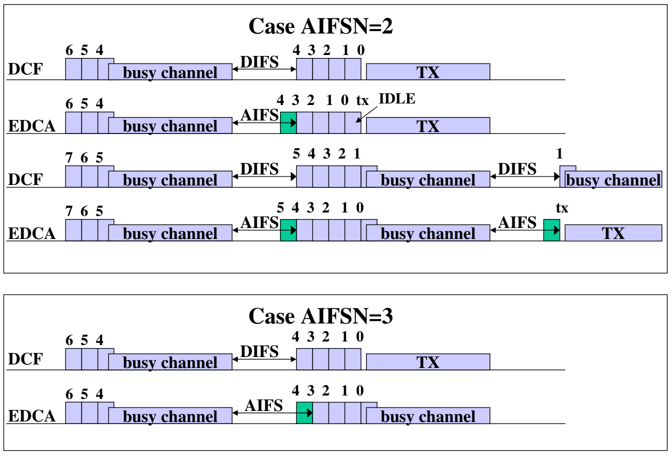

Fig. 4. Backoff counter management in EDCA and DCF

In EDCA, the backoff counter is also decremented at every idle slot-time and frozen during channel activity periods. But it is resumed one slot-time before the AIFS expiration. This means that, when the AIFS timer elapses, the backoff counter will result to be already decremented of one unit. Moreover, since a single MAC operation per-slot is permitted (backoff decrement or packet transmission, see [2], clause 9.9.1.3), when the counter decrements to 0, the station cannot transmit immediately, but it has to wait for a further backoff slot if the medium is idle, or for a further AIFS expiration if the medium is busy.

In order to understand how these different rules affect the channel access probability, refer to the example shown in figure 4. Let's first focus on the case AIFSN=2 (top figure), which corresponds to using an AIFS equal to the DCF DIFS. In the example, two stations encounter a busy channel period with the same backoff counter value. However, at the end of the channel activity, we see that the DCF station resumes its counter to a value equal to the frozen value (4 in the example), while the EDCA station resumes and decrements its counter. In the case of a single busy channel period encountered during the backoff decrement process, this difference will be compensated by the fact that the EDCA station will have to wait for an extra slot, i.e., unlike the DCF station, it will transmit in the slot following the one in which the backoff counter is decremented to 0 (this is the case illustrated in the two top diagrams of figure 4).

However, in the presence of several busy channel periods encountered during the backoff decrement process (which is very likely to happen in the presence of several competing stations), the EDCA station will gain a backoff counter decrement advantage for every encountered busy period with respect to the DCF station. This implies that, for an AIFSN equal to a DIFS, the EDCA station has an advantage over DCF. Indeed, figure 4 shows that there is also a second reason why, with same Inter Frame Space AIFSN=2, the EDCA gains priority over DCF stations. In fact, as shown in the figure, an EDCA station may actually transmit in the slot immediately following a busy channel period (it is sufficient that the busy channel period was encountered while the backoff counter was equal to 0 - last case in the figure). Conversely, a DCF station cannot de-freeze a backoff counter value equal to zero. Thus, the only case in which it can access the slot immediately following a busy period is when it extracts a new backoff counter, after a successful transmission, exactly equal to 0.

In order to synchronize the EDCA and DCF backoff decrements, it appears appropriate to set the AIFSN value equal to 3. In this case, as we can see in figure 4, although the EDCA station has an higher Inter Frame Space, after each busy slot the backoff evolution of the two target stations is the same. However, since the EDCA station has to wait for a further channel slot after the counter expiration, the access probabilities of the two stations does not coincide, since, for a given extraction, the EDCA station has always to wait for a slot more than the DCF station. However, this results in just a slightly higher access probability for the DCF station (loosely speaking, the EDCA station resembles the operation of a DCF station with a CW$_{min}$ value increased of just one unity).

B. Coexistence of EDCA AC_BE and legacy DCF stations

The throughput results shown in the earlier figure 3 show that, for the same contention window parameters, EDCA throughput performance are similar to that of legacy stations when the AIFSN parameter is set to the value 3 (i.e. the EDCA AC_BE Access Category, see table I), rather than to a legacy DIFS (i.e. AIFSN=2). The discussion carried out in the previous section has provided a qualitative justification.

Goal of figure 5 is to back-up the previous qualitative explanation with quantitative results. To this purpose, we have numbered slots according to the time elapsed after a busy channel period. The slot immediately following a DIFS is indexed as slot 0. Given the end of each busy period, the next transmission will occur after an integer number x of idle slots: we refer to x as the transmission slot. In the assumption of ideal channel conditions, a successful transmission occurs if, in a transmission slot, only one station transmits; otherwise a collision occurs. Figure 5 reports the probability distribution that

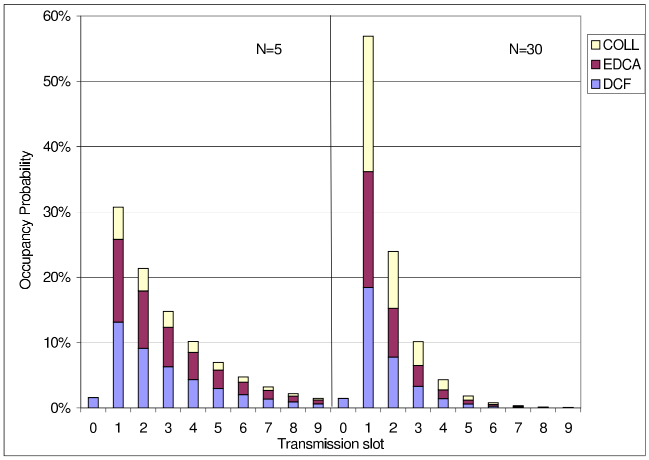

Fig. 5. Per-slot occupancy probability - AC_BE vs. DCF

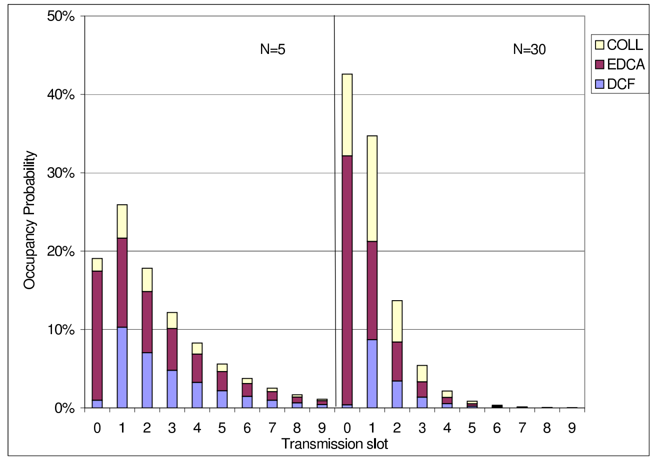

Fig. 6. Per-slot occupancy probability - AIFS differentiation

a transmission occurs at a given slot, for two different load scenarios: $N = 5$, i.e. 5 EDCA stations competing with 5 DCF stations, and $N = 30$. Only the first 10 slots are plotted, since most transmissions are originated after very few idle backoff slots. In addition, the figure further details in different colors the probability that a transmission occurring at a given slot results in a collision, in a success for an EDCA station, or in a success for a DCF station.

Figure 5 shows that DCF stations are the only ones that can transmit in the slot immediately following the last busy period. Also, it confirms that a transmission in slot 0 is always successful (as it is originated by a station that has just terminated a successful transmission). Indeed, a transmission in the slot immediately following a busy period is a rare event, since it requires that the station that has just experienced a successful transmission extracts a new backoff counter exactly equal to 0. The figure also shows that, in the slots with index greater than 0, DCF and EDCA stations experience almost the same success probability, with a negligible advantage for DCF. For example, in the case $N = 5$ a DCF success occurs, almost constantly through the various slot indexes, in about the 42.5% of the cases versus the 41% of EDCA, while for $N = 30$ these numbers reduce to, respectively, about 32.5% and 31.3% due to the increased probability of collision.

The fundamental conclusion is that *by using AIFSN=3, an EDCA station can be set to approximately operate as a legacy DCF station*. With reference to the proposed EDCA parameter settings reported in table I, we thus conclude that an EDCA station belonging to the Access Category AC_BE will experience similar performance than a legacy DCF station. The above quantitative analysis also justifies why DCF shows a slightly superior throughput performance over EDCA AC_BE, as found in the figure 3 under the case of AIFSN=3.

C. *AIFSN=2 and legacy DCF stations*

As shown in table I, AIFSN=2 is the minimal setting allowed for an EDCA station. The rationale is that both AIFSN=0 and AIFSN=1 are already reserved in the 802.11 standard for, respectively, the Short Inter Frame Space (SIFS), and for the Point Coordination Function InterFrame Space (PIFS).

However, as discussed above, the different mechanism employed in EDCA for decrementing the backoff counter suggests that, by using AIFSN=2 (i.e. AIFS=DIFS), an EDCA station is nevertheless expected to gain priority over a legacy DCF station. This was indeed shown in the previous figure 3, and is strikingly confirmed by figure 6, which, similarly to figure 5, reports the probability distribution that a transmission occurs at a given slot for the scenario of N DCF stations competing with N EDCA stations configured with AIFSN=2 and standard contention window parameters (CW$_{min}$=31 and CW$_{max}$=1023).

Figure 6 shows, for two different load conditions (N=5 and N=30), how the channel slots are occupied by the contending stations. From the figure, we see that the slot 0 results almost protected for EDCA stations, since it is rarely accessed by DCF stations. Channel slots with index higher than 0 are instead accessed by both the classes with comparable probability.

Figure 6 allows to draw a number of interesting considerations. First, the probability of collision in the protected slots (specifically, slot 0) is lower than in the other slots (e.g. for the case $N = 5$, a collision in slot 0 occurs only in about 8.5% of the cases, versus an average of 17% in the remaining slots, and these numbers for the case $N = 30$ become 24.5% versus 38.5%), due to the reduced number of competing stations. Second, and most

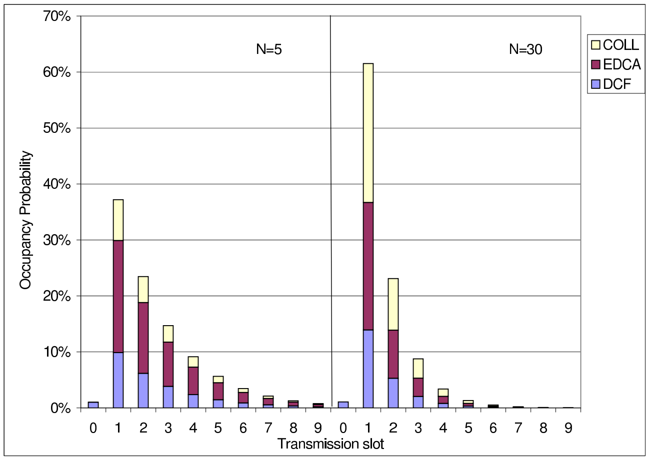

Fig. 7. Per-slot occupancy probability - CW$_{min}$ differentiation

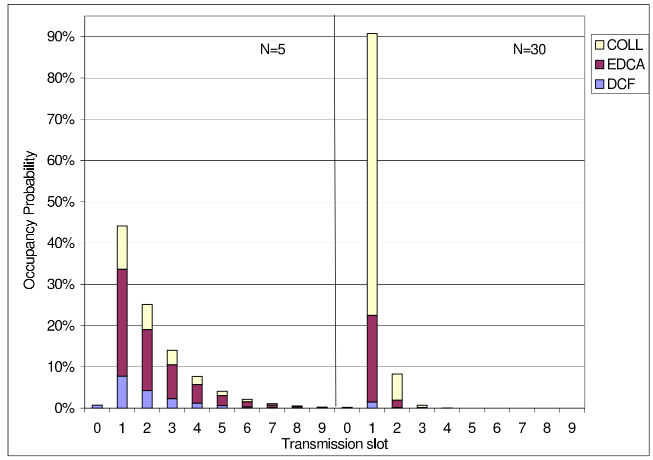

Fig. 8. Per-slot occupancy probability - CW$_{min}$ and CW$_{max}$ differentiation

interesting, as the network load increases, the probability of accessing low-indexed slots significantly increases. The reason is that the number of slots between two consecutive busy channel periods significantly reduces in high load. But this implies that a large amount of accesses occurs in slots 0 (more than 40% in the case of N=30, see figure 6), and thus are almost exclusively dedicated to EDCA stations, with a definite gain in terms of service differentiation effectiveness (as earlier shown in figure 3).

As a conclusion, *the usage of AIFSN=2 in EDCA (i.e. AIFS = DIFS) provides a significant priority of EDCA stations over legacy DCF stations*. This is an extremely important fact, as it allows to effectively deploy AIFS differentiation even when DCF stations share the same channel, and thus, apparently, there seems to be no room for AIFS levels intermediate between the Inter Frame spaces reserved by the standard (SIFS and PIFS), and the legacy DIFS.

D. Further remarks on AIFS vs CW$_{min}$ differentiation

The capability to dynamically adapt to network congestion, without significant performance impairments, makes AIFS differentiation an extremely effective approach. This is not the case when CW$_{min}$ differentiation is employed: as the number of stations in the network grows, too small CW$_{min}$ settings may lead to a dramatic increase of the probability of collision among frames. This is demonstrated in figure 7, which shows the probability that a transmission occurs at a given slot, and furthermore shows the probability that such a transmission results into a success or a collision. In figure 7, N DCF stations compete with N EDCA stations with parameters CW$_{min}$ $= 15$, CW$_{max}$ $= 1023$, and AIFSN=3.

By comparing figure 7 with figure 6, we can immediately observe the different effects of CW$_{min}$ and AIFS as differentiation parameters. In the case of AIFS differentiation, the number of transmission opportunities granted to different classes mainly differ because of the protected slots, while they are similar in the slots with index greater than 0. Conversely, in the case of CW$_{min}$ differentiation, the transmission opportunities are differentiated slot by slot. EDCA stations obtain more transmission grants because of the lower backoff expiration times. For example, in the left part of figure 7, for the case N=5, the EDCA successful accesses are almost the double of the DCF ones, in all the transmission slots which are accessed by both the classes. This happens because the DCF stations employ a CW$_{min}$ value which is the double of the EDCA one and the collision probability is small.

As the network congestion increases, it is more and more likely that EDCA stations experience collisions and double the contention window, so that the effects of the CW$_{min}$ differentiation is less effective. In fact, in the right part of the figure, we observe that for N=30 the increment of the collisions results in a reduction of the difference between EDCA and DCF successful transmissions, which are no more related by an exact 2:1 ratio (e.g. 13% DCF successes and 22% EDCA successes in the slot number 1). These considerations justify the curves behaviour in figure 1, where the main effect of the network load increment is the reduction of the EDCA throughput performance. Furthermore, the fiugure also shows that the probability to collide is significantly greater in the case of CW$_{min}$ differentiation.

Collisions can become dramatic in high load conditions, in the case of reduction of both the CW$_{min}$ and the CW$_{max}$ parameter, as suggested by the default values of table I for the access categories AC_VI and

AC_VO. Figure 8 shows the slot occupancy distribution for the case of N legacy DCF stations competing with N EDCA stations which exploit the default contention window parameters recommended for the AC_VI access category, namely CW_{min}=15 and CW_{max}=31. Different from table I, the AIFSN value has been set to 3, in order to specifically focus on just the effect of CW_{min} differentiation, and not on the joint effect of both mechanisms. The figure shows that the priority advantage of EDCA over DCF is provided through a significant increase of the probability that an EDCA station transmits in a low-indexed slot. However, the figure shows that a significant price to pay is a very high collision probability, resulting from the small adopted CW_{max}. This can be dramatic in high load conditions: as the number of stations grows, almost 80% of the transmission slots result in a collision. Note that with smaller CW_{min} and CW_{max} values (e.g. that recommended for AC_VO), the situation is much worse, and a significant collision probability may occur even with a small number of competing stations.

To conclude, the CW differentiation operation may lead to situations in which most of the channel slots are trashed by collisions. This can be considered as a fundamental inefficiency of the CW differentiation mechanism in high load conditions, which is unavoidable as long as small CW values have to be employed in order to allow prioritization over the legacy DCF stations.

IV. Conclusions

This paper has tackled the issue of the performance evaluation of EDCA service differentiation mechanisms. In comparison with the existing literature, we have put greater attention in analyzing the basic operation of the service differentiation mechanisms, not only in terms of high-level performance figures (such as throughput and delay), but also in terms of per-slot access probability distributions. This has allowed us to gain some additional understanding on the detailed operation of the EDCA mechanisms, and on the physical reasons behind their operation.

Our conclusion is that AIFS differentiation is a superior mechanism than CW_{min} differentiation, for a number of reasons. First, it does not trade-off service differentiation with aggregate performance impairment. Second, it is natively adaptive to network congestion. Third, even a single slot difference among AIFS values may result in a substantial difference in terms of performance.

A further contribute of this paper is the analysis of the coexistence between EDCA and legacy DCF stations. We have shown that the different backoff counter decrement mechanisms used in EDCA allows to gain, in practice, one extra slot to be used for AIFS differentiation: by setting the EDCA AIFS equal to the DCF DIFS (we recall that this is the minimum possible setting for the AIFS value, according to the present standard draft), EDCA traffic experiences a substantially higher access priority. Our results show that AIFS differentiation is effectively deployable in an hybrid EDCA/DCF scenario even if there seems to be a problem of lack of ”space” available.

References

[1] IEEE 802.11 WG, IEEE Std 802.11, 1999 edition. International standard for Information Technology. Telecommunications and information exchange between systems - Local and metropolitan area networks. Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, 1999.

[2] IEEE 802.11 WG, IEEE Std 802.11e/D10.0, Draft Amendment to STANDARD [for] Information Technology - Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Quality of Service (QoS) Enhancements, September 2004

[3] N. Ramos, D. Panigrahi, S. Dey, ”Quality of Service (QoS) Provisioning in 802.11e Networks: Challenges, Approaches, and Future Directions”, *IEEE Network*, this issue.

[4] Q. Ni, ”Performance analysis and enhancements for IEEE 802.11e wireless networks”, *IEEE Network*, this issue.

[5] S. Choi, J. del Prado, S. Shankar N, and S. Mangold, ”IEEE 802.11e Contention-Based Channel Access (EDCF) Performance Evaluation,” in *Proc. IEEE ICC’03*, Anchorage, Alaska, USA, May 2003

[6] D. He, C.Q. Shen, ”Simulation Study of IEEE 802.11e EDCF,” in *Proc. IEEE VTC Spring 2003*

[7] A. Lindgren, A. Almquist, O. Schelen, ”Quality of Service Schemes for IEEE 802.11 Wireless LANs - An Evaluation,” *ACM MONET*, Vol. 8, 2003, pp 223-235

[8] Y. Xiao, H. Li, and S. Choi, ”Protection and Guarantee for Voice and Video Traffic in IEEE 802.11e Wireless LANs,” *Proc. of IEEE INFOCOM 2004*

[9] J. Zhao, Z. Guo, Q. Zhang, and W. Zhu, ‘Performance Study of Service Differentiation in IEEE 802.11,” in *Proc. IEEE GLOBECOM’2002*

[10] Y. Xiao, ”An Analysis for Differentiated Services in IEEE 802.11 and IEEE 802.11e Wireless LANs,” in *Proc. IEEE ICDCS’2004*, March 2004

[11] Martin Heusse, Franck Rousseau, Gilles Berger-Sabbatel, Andrzej Duda, ”Performance Anomaly of 802.11b”, *Proc. of IEEE INFOCOM 2003*

[12] G. Bianchi, ”Performance analysis of the IEEE 802.11 distributed coordination function,” *IEEE JSAC*, Vol.18, No.3, Mar 2000, pp. 535-47.