

Performance Evaluation of Differentiated Access Mechanisms Effectiveness in 802.11 Networks

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

^(*) 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

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

Abstract—The IEEE 802.11e draft specification aims to extend the original 802.11 MAC protocol by introducing priority mechanisms able to manage bandwidth and resource allocation according to the QoS needs of real-time applications. Different strategies based on MAC parameter diversifications, such as contention window limits, contention window updating factor and silence monitoring time, can be pursued in order to provide service differentiation, also in the case of distributed access. In this paper, we investigate on the behaviour of each differentiation possibility under different load conditions and traffic requirements. Our results show that the most powerful mechanisms which provide service differentiation are based on inter frame space (*IFS*) and minimum contention window (CW_{min}) diversification. However, these mechanisms have a different behaviour, since CW_{min} different settings affect not only access priorities, but also the overall system performance. We conclude that the best differentiation policy should be based on a joined use of *IFS* and CW_{min} differentiation, using the first parameter to adjust access probabilities, and the second parameter to optimize the resource utilization in the network¹.

I. INTRODUCTION

The IEEE 802.11 WLAN [1] standard is expected to become a widespread technology for the deployment of multimedia services over hot-spots and home-networking environments. Since 2000, over 40 million of 802.11 devices have been deployed worldwide and a more and more increasing number of users, after experiencing the convenience of wireless connections through 802.11 cards, asks for support of time-sensitive (voice and video) applications they usually run over wired networks. In order to satisfy these new market needs, the 802.11 task group "E" is currently working to enable the QoS support within the 802.11 framework. The extensions introduced regard both the distributed (DCF) and the centralized (PCF) access schemes, and are mainly based on the non-uniform settings of backoff and channel monitoring parameters. Although the 802.11e standard is still in draft version (the final standard is expected for the end of 2004/beginning 2005), in the meantime, the large amount of work carried out by the scientific community has already pointed out that the current 802.11e proposal presents a number of important issues to be solved. Because of the earlier and faster success obtained by the distributed access mechanism (EDCA) over the centralized access one (HCCA), most of these studies have been oriented on the performance evaluation of the EDCA proposal [5], [6], [7]. From this research activity has emerged that, although the EDCA [2] specification proposes a large number of technical solutions, it is still unclear both what is the performance effectiveness of these new service differentiation mechanisms, individually taken, and what is the robustness of the performance achieved by given

parameter settings with respect to traffic parameters (e.g. number of contending stations). A very comprehensive overview of the various complementary as well as mutually alternative mechanisms for service differentiation in 802.11 can be found in the recent technical report [3], which also addresses the performance evaluation of a subset of the considered mechanisms. Another comparative performance evaluation, although for a limited number of proposals (four) can be found in [4]. Some analytical models [9], [10], [6] have been developed in order to evaluate system performance (in terms of throughput and delay repartitions among the classes) as a function of the MAC settings and of the number of stations for each class. It has been proposed to use such results for analysis and admission control purposes [6]. However, little effort has been done for the synthesis and the service differentiation engineerization. In this paper we focus on this last issue and, in particular, on the problem of service differentiation, while maintaining the best possible resource utilization in the network. We show that a powerful mechanism is the combined use of various differentiation strategies (namely, CW_{min} and *IFS*). The rest of the paper is organized as follows: in section II, we briefly describe the differentiation mechanisms present in the EDCA proposal, in section III the effect of each differentiation strategy is explained and quantified via simulation results in different load scenarios, while in section IV some conclusions are drawn.

II. ENHANCED DISTRIBUTED CHANNEL ACCESS

The Enhanced Distributed Channel Access (EDCA) proposal enhances the current DCF access mechanism of the IEEE 802.11 standard in order to provide service differentiation using a distributed control approach. It is devised to differentiate the channel access probability among different traffic sources. The mechanism is designed to manage 8 different traffic priorities. The packet priority represents the quality of service defined at higher layers. Packets arriving to the MAC layer (Mac Service Data Units MSDUs) are then 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 following the same rules as the legacy DCF, i.e. wait until the channel is idle for a given amount of time (*DIFS* in standard DCF), and then access/retry following exponential backoff rules. Every AC corresponds to a separate and independent queue at the MAC layer, and behaves as a single contending DCF entity. When more than one AC at the same station simultaneously expire their own backoff counter, a virtual collision occurs among them. In this case, an internal scheduler resolves the contention by selecting the highest priority packet among the colliding ones for the actual transmission on the radio channel. Basically, the EDCA mechanism defines both

¹The work reported in this paper was partially supported by the Italian Research Program PRIMO.

AC	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

different silence monitoring times and different backoff parameter values in order to configure the Arbitration Inter Frame Space and the minimum/maximum contention window values for each of the four supported AC (respectively the $AIFS[AC]$ value, the $CW_{min}[AC]$ value and the $CW_{max}[AC]$ value). In addition, previous 802.11e drafts also included other differentiation mechanisms which aren't anymore considered since the version 5.0 of the 802.11e Draft proposal. However, in the following, we also discuss these further differentiation possibilities. The $AIFS[AC]$ values differ each one for an integer number of backoff slots. In particular, the $AIFS[AC]$ value is equal to $AIFSN[AC] \cdot aSlotTime + aSIFSTime$, where the $AIFSN[AC]$ value is an integer number greater than 1 for normal stations and greater than 0 for APs. The initial and final contention window values depend on the specific AC ($CW_{min}[AC]$ and $CW_{max}[AC]$). Table I reports the proposed EDCA default $AIFSN$, CW_{min} and CW_{max} settings of each of the four ACs.

The backoff time is calculated as in legacy DCF, by drawing a random number in the interval $[0, CW - 1]$, whereas the CW value is determined, and updated after each failed transmission, by using the basic DCF rules. Nonetheless, in previous 802.11e draft proposals, another mean to achieve service differentiation had been individuated in the Persistence Factor ($PF[AC]$), another AC dependent parameter which affects the CW updating rule.

As in legacy DCF, when the medium is determined busy before the backoff counter is zero, the station freezes its count down until the medium is sensed idle for an $AIFS$ again, but differently from that, the backoff counter is reduced by one at the beginning of the last slot interval of the $AIFS$ period.

In each beacon frame, the AP broadcasts the values of the differentiation parameters chosen for each AC. In fact, they can be dynamically adapted according to the network conditions. Once a station wins the contention and starts its transmission grant, EDCA also specifies new channel utilization operations based on the concept of transmission opportunity (TXOP), which represents a time interval in which the station is authorized to hold the channel.

III. DIFFERENTIATION MECHANISMS BEHAVIOUR

As we briefly described in previous section, several solutions have been proposed to provide *access* differentiation in a distributed manner. Obviously, these operations do not allow to offer any form of guarantee to the users. Service priority has to be intended as an increment of the probability to win the contention for the access to the common channel. While in DCF the winner of each contention gains a packet transmission opportunity, in EDCA the winner of a contention obtains the channel for a given period of time (TXOP). Roughly speaking, the effect of MAC parameter differentiations is the loss of access fairness. Note that service differentiation problem is much more complex than a simple bandwidth repartition among the classes, since different traffic requirements (such as maximum tolerable delay or delay jitter) have to be taken into account. In fact, differentiated bandwidth repartition can be easily obtained even in standard DCF by giving different maximum payload size to different stations (or, equivalently in EDCA, by giving different TXOP), but this operation does not allow to differentiate the service delays.

Thus, we want to clarify what type and what amount of service differentiation can be obtained by giving, for each station, different settings to the following MAC parameters: CW_{min} , CW_{max} , PF , $AIFS$. To this purpose, we developed an object-oriented event-driven C++ simulator, in which we implemented all the described differentiation access features. In the simulator, differentiation mechanisms can be separately or conjunctly activated, while each station has a single traffic flow. We compare the performance obtained by QoS stations which adopt different access parameters. For each differentiation parameter, we run simulations for the basic access (without using RTS/CTS) varying the parameter settings and the traffic conditions, in order to identify equivalences among different mechanisms. For example, we investigate if the effect of setting different $AIFS$ is somehow equivalent to different CW_{min} settings. We consider two different types of traffic sources. Data transfer sources are modelled with saturated queues at the MAC level, while real-time sources are modelled with poisson arrivals. Setting values used in simulation are equal to those defined in the standard [1] for the 802.11b PHY.

A. Homogeneous sources

Consider preliminarily the effect of MAC parameter differentiation on resource assignment, when all the stations are involved in data transfer. We assume that each station has always a packet to send (i.e. transmission queues are never empty) and that sources are homogeneous in terms of packet length (1024 bytes) and data rate (11 Mbps, unless otherwise specified). In this scenario, when the radio access is managed by basic DCF, resource repartition among the stations is fair and, averagely, each station wins the same number of contentions. The winner of the contention gains a packet transmission opportunity. Since the packet collision probability and the packet length are constant for each station, probabilistically there is no difference in the amount of data transmitted with success by all the stations. Let now see how this fairness can be lost. In particular, we want to divide the contending stations into two service classes: high priority (HP) and low priority stations (LP).

1) *Maximum Contention Window*: When all the stations are in contending state, the access to the channel is granted to the station which extracts the lower backoff value. In order to advantage some stations to others, we can specify different CW_{max} values. In case of consecutive collisions, stations with lower CW_{max} have a greater probability to extract the minimum backoff and then to access the channel. Figure 1 shows the effect of CW_{max} differentiation among two service classes. In the simulations, we set an equal number N of LP and HP stations and we vary the CW_{max} value assigned to HP stations. LP stations are assumed to adopt the standard $CW_{max} = 1024$ value, while HP stations use lower CW_{max} values. The figure plots the LP, HP and total throughput versus the number N of stations for each service class, varying the HP CW_{max} values. From the figure, it can be noticed that throughput differentiation is not very significant in every network configuration, and especially when N is low. Only for CW_{max} values lower than 64 and for N greater than 15, LP throughput and HP throughput are visibly differentiated, but in these cases there are also great inefficiencies in the channel utilization (as shown by the total curve). In fact, this form of differentiation operates only after consecutive collisions (sufficient to update HP CW to its maximum value) occur. These events are more and more frequent as the value of CW_{max} is low and as the number of stations increases. But frequent collisions (needed for service differentiation) waste a large amount of resources, and then reduce the total channel utilization.

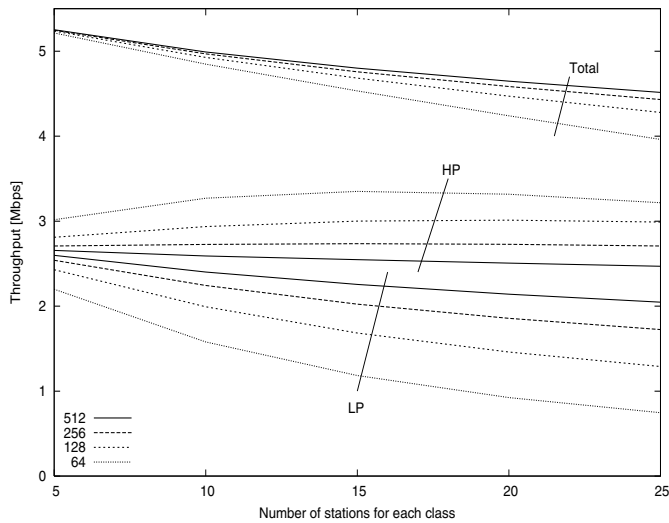


Fig. 1. Throughput vs. the number of stations for each class varying the HP CW_{max} value

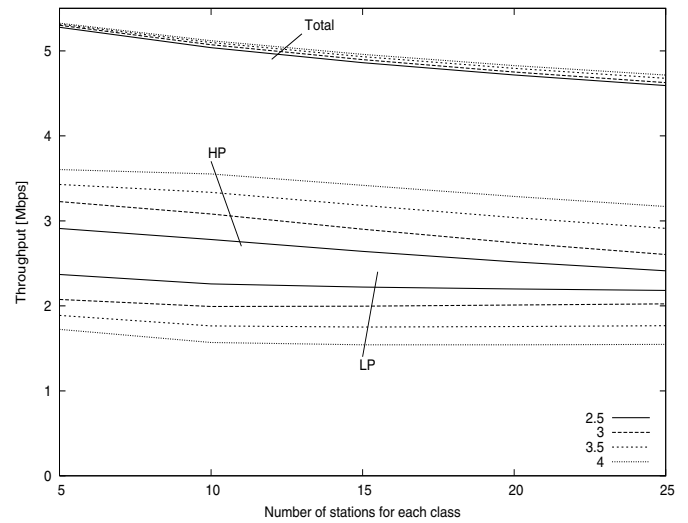


Fig. 3. Throughput vs. the number of stations for each class varying the LP PF value

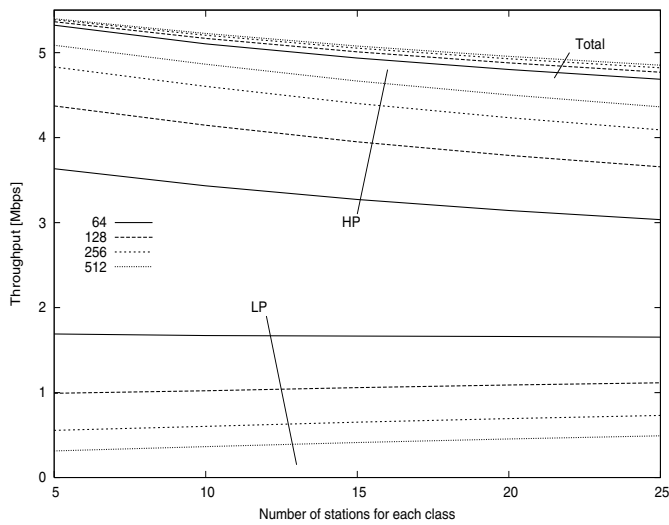


Fig. 2. Throughput vs. the number of stations for each class varying the LP CW_{min} value

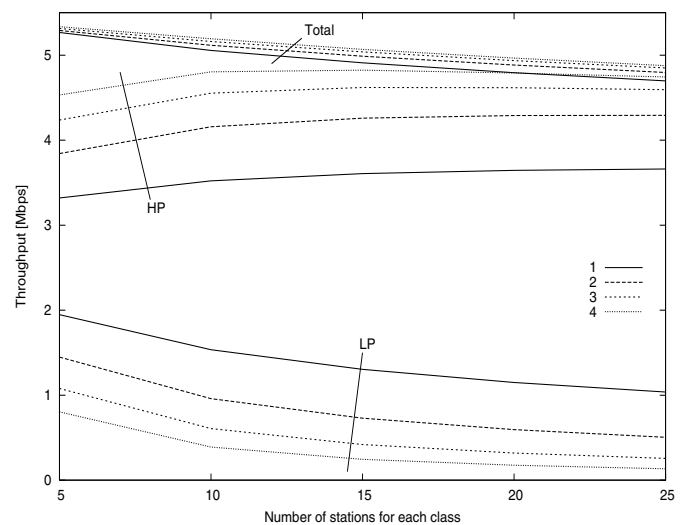


Fig. 4. Throughput vs. the number of stations for each class varying the LP IFS value

2) *Persistent Factor*: After each collision, the contention window is updated according to the law $CW_{new} = CW_{old} * PF$ until the maximum value is reached. Another strategy to give priority is the adoption of a persistent factor PF greater than 2 for LP stations². Again, a higher CW value corresponds to a lower probability to extract low backoff values and to access the channel. Figure 3 plots the LP, HP and total throughput versus the number N of stations for each service class, varying the LP PF values. From the figure, we see that service differentiation is evident, but we cannot choose an opportune PF value which corresponds to a desired HP/LP throughput ratio. In other words, it is not possible to force LP throughput to zero and to give all the resources to HP stations by simply choosing a high PF value. This phenomenon is due to the fact that also this form of differentiation operates after collisions. Although it is sufficient a single collision event to begin the differentiation process, LP stations

²Equivalently, we can choose PF values lower than 2 for HP stations. System behaviour tends to be similar to the case discussed in previous section with CW_{max} close to 32

contend equally with HP stations after packet successful transmission and then receive a given amount of bandwidth which does not depend on PF .

3) *Minimum Contention Window*: If we want to differentiate access probabilities before collisions, a possible strategy is the adoption of different CW_{min} values for HP and LP stations. Again the idea is the increment of the extraction probability of the minimum backoff value for HP stations. In this case, we assume that HP stations are legacy stations ($CW_{min}=32$) and run simulations varying the CW_{min} settings for LP stations. Figure 2 plots the LP, HP and total throughput versus the number N of stations for each service class, varying the LP CW_{min} values. Note that in this case LP throughput approaches zero as the value of CW_{min} increases. Thus, for a given load scenario and a given desired throughput ratio it is possible to choose an opportune setting for LP CW_{min} .

4) *Inter Frame Space*: This form of service differentiation works on channel monitoring parameters rather than on backoff parameters. Basically, different inter-frames spaces are already used in DCF

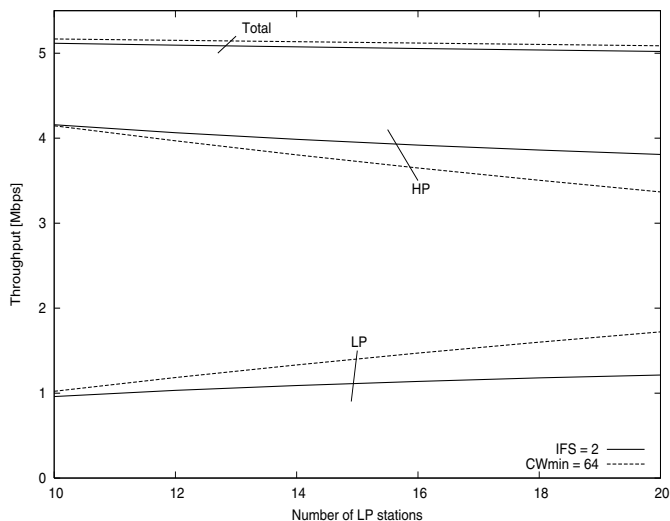


Fig. 5. Throughput vs. the number of LP stations for different IFS and CW_{min} values

to give priority to ACK transmissions ($SIFS$) or to contention free periods ($PIFS$) on data packets. The idea is to extend this mechanism using different $DIFS$ values for each service class. In particular, for each Access Category, we consider an $AIFS[AC]$ value equal to a $DIFS$ time plus a given AC-dependent number of time-slots. In our simulations, we assume that HP stations adopt a monitoring time equal to a $DIFS$ and that LP stations adopt higher silence monitoring values. Figure 4 plots the LP, HP and total throughput versus the number N of stations for each service class, varying the LP additive time-slot number IFS . The behaviour of the curves is quite similar to the one illustrated for the CW_{min} case. By comparing this figure with the previous one, note that only one slot of IFS differentiation is sufficient to obtain almost the same throughput ratio correspondent to a doubled LP CW_{min} . Even if both the mechanisms allow a full control on service differentiation, some differences are evident as the network load conditions change. We note that, as the network congestion increases, IFS differentiation

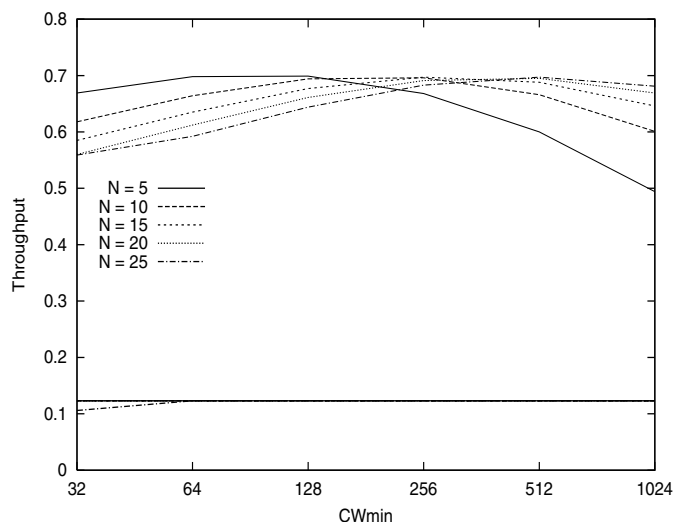


Fig. 6. Throughput vs. CW_{min} settings varying the number of data sources

appears more robust, guaranteeing a reliable resource allocation for HP stations, whereas CW_{min} differentiation tends to penalize them. This phenomenon is not desirable, since performance degradations due to the network congestion should be attributed first to LP stations. Considering the IFS case (fig 4), performance degradation due to an increasing number of active stations, is mostly attributed to LP stations, while HP throughput is close to saturation conditions and slightly degrades as the number of stations grows. Conversely, in the CW_{min} case, the total throughput decrement due to congestion is mainly suffered by HP stations.

Another important difference is illustrated in figure 5, in which we have a different number of LP and HP stations. In the figure, we run simulation varying the number of LP stations and maintaining the number of HP stations equal to 10. Two groups of curves are plotted for two different differentiation possibilities for LP stations (namely, IFS value equal to 2 or CW_{min} value equal to 64). It is evident that IFS differentiation division of channel resources among the classes is less sensitive to the number of LP stations, since the slope of their throughput curve is lower.

B. Heterogenous sources

We now consider the problem of resource sharing among heterogeneous sources. In particular, we assume that contending stations are involved into two different types of applications: data transfer or low rate delay-sensitive traffic. Stations involved in data transfer operate in saturation conditions, while stations involved in low rate traffic generate packets with exponentially distributed inter-arrival times. We assume that packet length is still homogeneous among all the stations (and equal to 1024 bytes), while the average generation rate for low rate sources is equal to 3 packets/s. The number of such sources is constant and equal to 5, while the number N of data sources is a simulation parameter. Different strategies can be adopted in order to give priority to delay-sensitive traffic (HP stations). We focus on CW_{min} and $AIFS$ settings of data stations (LP stations) since, as we saw in previous section, these parameters are more effective. Figures 6 and 7 show the HP and LP throughput for this simulation scenario and 1 Mbps of data rate. HP throughput is equal to the HP offered load (namely, 122880 bit/s) for almost all the configurations; only in absence of service differentiation ($IFS = 0$ or $CW_{min} = 32$) it

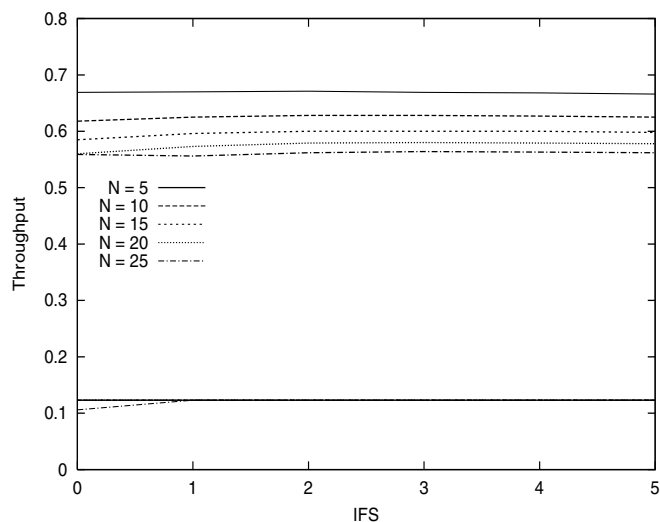


Fig. 7. Throughput vs. IFS differentiation settings varying the number of data sources

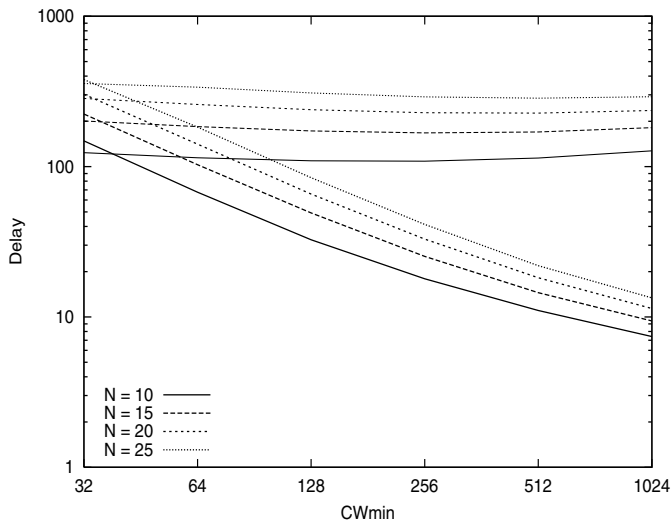


Fig. 8. Service Delay vs. CW_{min} settings varying the number of data sources

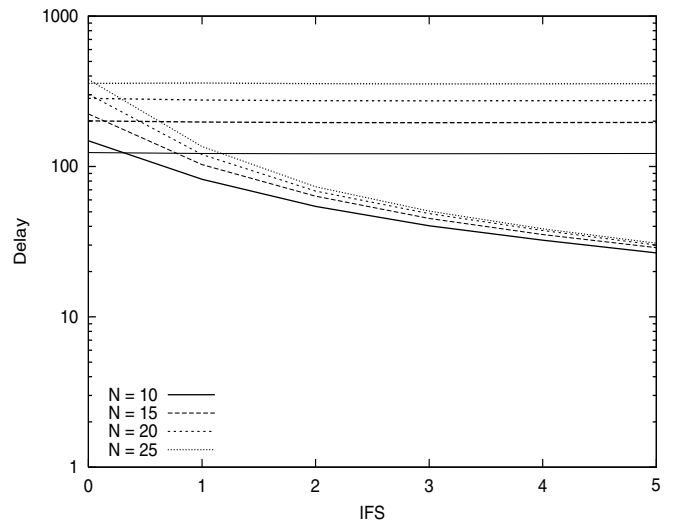


Fig. 9. Service Delay vs. IFS differentiation settings varying the number of data sources

results lower when the number of data sources is equal to 25. In fact, as the number of contending data sources increases, when HP stations generate new packets, the probability to win the contention gets lower and lower. Conversely, it is sufficient a low IFS or CW_{min} differentiation to give the required bandwidth to HP stations, even in heavy data load conditions. If we further increase the differentiation of the CW_{min} or IFS values among the classes, HP performance improve in terms of service delays as shown in figures 8 and 9. This implies that, service differentiation parameters can be chosen according to the HP delay requirements. Note that in the IFS case, delay performance for HP stations are less sensitive to the number of data stations N (curves are less spread) and that LP CW_{min} doubling has more effect on HP delay reduction than IFS increasing of one time-slot. Of course, not every requirement can be satisfied since, even completely excluding LP stations from the channel access, HP stations are in competition among them.

About data sources, from the figures we note that throughput has a very different behaviour in the case of CW_{min} or IFS differentiation. In the former case it strongly depends on the CW_{min} setting, while in the latter case it is almost independent on the IFS value. In fact, data throughput corresponds to the saturation throughput of N stations which share the resources exceeding to HP stations. Hence, as shown in [8], this throughput presents an optimum value for a given CW_{min} value which depends on N , while it is almost independent on the channel monitoring times. This implies that, according to the CW_{min} value chosen to satisfy delay requirements for HP stations, data stations performance can vary significantly and can be far from optimum value. Thus, we argue to combine the use of CW_{min} and IFS differentiation: CW_{min} value has to be set in order to maximize data throughput, and opportune IFS differentiation has to be introduced to satisfy HP delay requirements.

IV. CONCLUSIONS

In our study we present an overview of the new QoS supporting features of EDCA. The performance of the technical solutions in phase of debate (and in particular of minimum contention window and silence monitoring time differentiation) for QoS support are discussed via simulation results. Although each solution constitutes a possible mean for QoS support in WLANs for a wide variety of applications,

simulations show that the analyzed mechanism have a very dissimilar behaviour varying the traffic conditions in the network. Effective parameter settings to cope with the optimization of network resources are investigated. We also remarked that all forms of priority do not give absolute guarantees about throughput and delay, but only relative differentiation among service classes. Further investigations will regard the tuning of the optimal differentiation parameter settings in the case of joined CW_{min}/IFS differentiation.

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/D8.0, 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, February 2004
- [3] Qiang Ni, Lamia Rhomdani, Thierry Turletti and Imad Aad, QoS Issues and Enhancements for IEEE 802.11 WLAN, INRIA technical report RR4612, November 2002, <http://www.inria.fr/tr-4612.html>.
- [4] Anders Lindgren, Andreas Almquist, Olov Schel, Evaluation of Quality of Service Schemes for IEEE 802.11 Wireless LANs, LCN 2001.
- [5] I. Aad, C. Castelluccia, Priorities in WLANs, Computer Networks (Elsevier), vol. 41, 2003, pp. 505-526.
- [6] A. Veres, Campbell, A.T. Barry, M. and L-H. Sun, "Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control", IEEE Journal of Selected Areas in Communications (JSAC), Special Issue on Mobility and Resource Management in Next-Generation Wireless Systems, Vol. 19, No. 10, pp. 2094-2104, October 2001.
- [7] Sunghyun Choi, Javier del Prado, Stefan Mangold, and Sai Shankar, IEEE 802.11e Contention-Based Channel Access (EDCF) Performance Evaluation, Proc. IEEE ICC03, Anchorage, Alaska, USA, May 2003.
- [8] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function", IEEE JSAC, Vol.18, No.3, Mar 2000, pp 535-47.
- [9] G. Bianchi, I. Tinnirello "Analysis of priority mechanisms based on differentiated Inter Frame Spacing in CSMA-CA", IEEE VTC 2003 Fall, Oct 2003, Orlando.
- [10] J. Zhao, Z. Guo, Q. Zhang, W. Zhu, "Performance study of MAC for service differentiation in IEEE 802.11", GLOBECOM 2002. IEEE, Volume: 1, Nov 17-21, 2002