Design and Performance Evaluation of a Distributed Contention Control (DCC) Mechanism for IEEE 802.11 Wireless Local Area Networks[1]

Luciano Bononi, Marco Conti(*), Lorenzo Donatiello

Dipartimento di Scienze dell'Informazione
Università di Bologna
Mura Anteo Zamboni, 7
40127, Bologna, Italy
e-mail: {bononi, donat}@cs.unibo.it

(*) CNR Istituto CNUCE
Via S.Maria, 36
56100, Pisa, Italy
e-mail: M.Conti@cnuce.cnr.it

This paper focuses on the design and the performance evaluation of a new mechanism, named *Distributed Contention Control (DCC)*, for the adaptive contention reduction in LAN networks that utilize random access MAC protocols. The proposed mechanism could be executed on the top of a pre-existent access protocol, with no additional overhead introduced. Specifically, we consider the IEEE 802.11 wireless LAN (WLAN). The DCC mechanism requires a simple and rough estimate of the contention level, and this can be achieved by estimating any parameter, directly connected with the amount of contention on the shared channel. The main characteristics of the proposed mechanism are represented by its simplicity, integration with the Standard, complete distribution, absence of overheads, and prompt reaction to changes in the network congestion. The protocol automatically adapts to the network congestion, by monitoring the channel contention level, through the estimation of the contention parameter. In the paper we show that the information needed for the contention estimation is already available to a 802.11 station, with no additional costs. Simulation experiments to evaluate the performance of an IEEE 802.11 WLAN, with and without the DCC mechanism, have been carried out. Results confirmed the effectiveness of the DCC mechanism in improving the performance, stability, and congestion reaction of the IEEE 802.11 access scheme. The DCC mechanism also provides a simple way to implement a distributed priority mechanism.

Key Words: wireless LAN, IEEE 802.11, congestion control, performance evaluation, simulation

[1] This work has been partially supported by Murst and CNR.

1. INTRODUCTION

This paper proposes a *Distributed Contention Control (DCC)* mechanism for improving the performance of the IEEE 802.11 access scheme for wireless LANs [8]. The IEEE 802.11 access scheme is based on two access methods: *Distributed Coordination Function (DCF)* for asynchronous, contention-based, distributed accesses to the channel, and *Point Coordination Function (PCF)* for centralized, contention-free accesses. We will concentrate our analysis over DCF which, due to distributed random access characteristics, may be significantly affected by the congestion problem.

The DCF method adopts a distributed, contention-based access scheme belonging to the class of the CSMA/CA MAC protocols [5, 8, 12]. Collisions occur when two or more users simultaneously try to utilize the resource. When a collision occurs, the resolution of the contention is obtained with a variable time-spreading of the users' accesses. Hence, this system is subject to a resource wastage caused both from collisions and from the resource idle periods introduced by the accesses' time-spreading. As the reduction of the idle periods generally produces an increase in the number of collisions, to maximize the resource utilization the CSMA protocol should balance these two costs [4, 9]. Since these costs change dynamically, depending on the network load, it is evident the need to make the CSMA protocol adaptive to the congestion level in the system [5, 9, 13]. The DCF adaptive behavior is obtained, independently by each station, by means of a *backoff protocol* that exploits the station history regarding successes or collisions [13]. Specifically, each user is not assumed to have any knowledge about other users' successes or collisions, or even about the number of users in the system. Each station transmits on the channel within a random self-defined amount of time, whose mean size depends on the number of collisions already experienced by that station while transmitting a given frame. For these protocols, when the number of users grows, or even in bursty-arrival scenarios, a low resource utilization is often obtained, due to high-collision rates [5, 13]. This fact seems difficult to avoid because the system's reaction to high-contention conditions is based on collisions. Examples of these channel allocation problems arise in the context of CSMA/CD multiple access protocols for LANs such as Ethernet and IEEE 802.3, or in the IEEE 802.11 MAC protocol for WLANs [8]. Specifically, it is known that for CSMA/CD MAC protocols, given a great number of users, the communication capability of the network could collapse [5, 13].

Several authors have investigated the enhancement of the IEEE 802.11 DCF MAC protocol to increase its performance when it is used in WLANs. In [6, 7], via a performance analysis, it has been studied the tuning of the Standard's parameters. In [20], given the *Binary Exponential Backoff* scheme [8, 11] adopted by the Standard, solutions have been proposed for achieving a more uniform distribution of the accesses.

The backoff protocols have been extended by studying the information which can be obtained by observing the system's parameters [10, 15, 19]. Some authors try to approximate the knowledge about the number of users involved in the channel accesses by means of the analysis of the past history of the system. For example, in the case of the IEEE 802.11 DCF MAC protocol, it has been proposed to make the contention reduction adaptive to the estimated number of users in the system [1, 4]. It is worth observing that the *number of users* estimate could result expensive, difficult to obtain and subject to significant errors, especially in high contention situations.

In this work, we propose a new contention-reduction mechanism, named DCC, simply based on a rough and easy to obtain estimate of the contention level, named $Slot\_Utilization$.

Our mechanism is adaptive to the contention level of the system, by exploiting the proposed contention-level estimate. Hence, the contention reaction performed does not depend on the collisions' occurrence, rather contributes to prevent them. Moreover, the contention reduction is achieved, in a distributed way (i.e. independently by each station), at zero cost and with simple distributed management. It is worth noting that, with the DCC mechanism (implemented on top), it results possible to extend even a simple 1-persistent scheduling protocol (e.g. Ethernet), whatever the backoff function adopted, obtaining a powerful and adaptive mechanism for the contention control in accessing the channel.

The work is organized as follows: in Section 2 we present the IEEE 802.11 protocol and we discuss the problem related to the congestion reaction; in Section 3 we present the DCC mechanism and we discuss its implementation; in Section 4 we study some characteristics of the system obtained; in Section 5 we present simulation results of the proposed mechanism, implemented on top the IEEE 802.11 DCF protocol; in Section 6 we introduce a priority-based access scheme, and we present some related simulation results; in Section 7 conclusions and future researches are drawn.

2. IEEE 802.11 STANDARD DCF FOR WIRELESS LANS

For the detailed explanation of the IEEE 802.11 Standard, and in particular of the Standard DCF access method, we address interested readers to [6, 7, 8, 16].

The 802.11 DCF is based on a CSMA/CA MAC protocol which requires every station to perform a Carrier Sensing activity to determine the current state of the channel (idle or busy). This allows each station to evaluate the opportunity to start a transmission without interfering with any other ongoing transmission. If the medium is found to be busy, each station defers the transmission. Whenever the channel becomes idle for at least a *Distributed InterFrame Space* time interval $(DIFS)$, the station (re)starts its Basic Access mechanism. To avoid collisions after a DIFS idle period, a *Collision Avoidance* mechanism is needed. The Collision Avoidance mechanism adopted in the 802.11 Standard is based on a *Binary Exponential Backoff* scheme. The objectives of the backoff scheme are: *i)* a distribution (the most uniform as possible) of the transmission attempts over a variable-size time window, and *ii)* small mean access delay under light load conditions. When the channel is idle, time is measured in units of constant length indicated as *slots* in the following. The Binary Exponential Backoff scheme is implemented by each station by means of a parameter, named *Backoff_Counter*. This counter maintains the number of empty slots a station must observe on the channel before performing its own transmission attempt. When a tagged station has a packet to transmit, it initializes the *Backoff_Counter* according to the following expression [8]:

$$Backoff\_Counter(CW\_Size) = int(Rnd() * CW\_Size)$$

where $Rnd()$ is a function returning pseudo-random numbers uniformly distributed in [0..1], and CW_Size is a local parameter whose value (expressed in slots) represents the size of the time window into which the transmission attempt will occur. The $Backoff\_Counter$ is decreased as long as a slot time is sensed as idle, it is frozen when a transmission is detected, and reactivated after the channel is sensed as idle for at least a DIFS. As soon as the $Backoff\_Counter$ reaches the value $Zero$ the station can transmit its own frame.

If a station transmission attempt generates a collision[1], to reduce the collision risk, the station doubles the CW_Size parameter for scheduling the retransmission attempt.

The *Binary Exponential Backoff* is then characterized by the expression which gives the dependency of the CW_Size parameter by the number of transmission attempts (Num_Att) already performed for a given frame, including the current one. In [8] it is defined that the first transmission attempt for a given frame is to be performed with CW_Size equal to the minimum value CW_Size_min, i.e. it is assumed a low-contention condition. After each unsuccessful (re)transmission of the same frame, the station doubles CW_Size until it reaches the maximum value defined in the Standard, i.e. CW_Size_MAX, as follows:

$$CW\_Size(Num\_Att) = min(CW\_Size\_MAX, CW\_Size\_min * 2^{Num\_Att-1})$$

The increase of the CW_Size parameter value, after a collision, is the reaction that the 802.11 Standard provides to make the access mechanism adaptive to channel conditions.

2.1. Congestion reaction analysis

By analyzing via simulation the behavior of the 802.11 DCF mechanism, under various contention levels (i.e. the number of active stations with continuous transmission requirements), some problems could be identified, see Fig. 1. Specifically, Fig. 1 shows the channel utilization of a Standard 802.11 system running in DCF mode, with respect to the contention level[2]. The results show that the channel utilization is negatively influenced by the increase of the contention level. This can be explained by observing that, according to

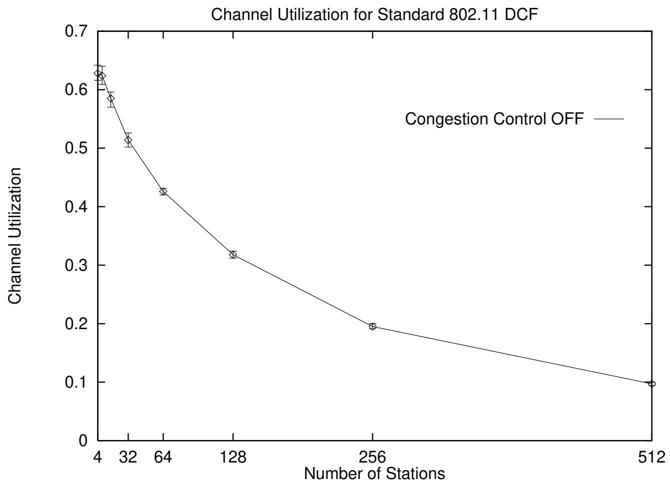

FIG. 1. Channel utilization of Standard 802.11 DCF

the 802.11 Standard, when a station has a new frame to transmit, it selects the initial size of the Contention Window (CW_Size) by assuming a low level of congestion in the system. This choice enables an "on-the-fly" transmission when the load is light. Unfortunately, this represents a problem in bursty-arrival scenarios, and in congested systems, because it introduces a concentration of accesses in a small time window, and hence it causes a high

[1] A collision is assumed whenever the Ack from the receiver is missing [8].
[2] The parameters adopted in the simulation are presented in Table 2 of Section 5.

collision probability. According to the Standard, when the network is highly congested, a station reacts to it only on the basis of the unsuccessful transmission attempts already experienced for a given frame. Every station performs its attempts blindly, with a late collision reaction performed by increasing the CW_Size for successive retransmission attempts. In addition, several retransmissions of the same frame introduce a further increase in the network congestion. Moreover, Collision Detection (CD) results not practical in wireless channels, and this makes the cost of each collision proportional to the maximum size among colliding frames.

To summarize, the information used to determine the contention level in the 802.11 Standard has two negative aspects: *i)* a station requires several collisions to be aware of a high-contention level, and *ii)* each collision does not provide a significant indication of the actual contention level, due to stochastic variability in the slot selection[3].

3. THE DCC MECHANISM

The reasons explained in the previous section, motivated us to study the possibility to improve the performance of a CSMA/CA access scheme, by exploiting early and meaningful information concerning the actual channel congestion level. Our idea is based on an estimate of the channel's congestion level, obtained by measuring the utilization rate of the slots $(Slot\_Utilization)$ observed on the channel by each station. It is easy to observe that the $Slot\_Utilization$ could assume the role of an early and significant Contention Level indication. In Section 3.1 we discuss how we compute the $Slot\_Utilization$ index to estimate the network congestion level. This index is used in Section 3.2 to implement a distributed contention-control mechanism, named DCC, for accessing the shared channel.

3.1. The Slot Utilization: an early estimate of the contention level

The *Slot Utilization* $(Slot\_Utilization)$ estimate we propose to utilize is simple, with zero costs, and no overheads on the protocol performance. All the information needed for the $Slot\_Utilization$ estimate is already available in an IEEE 802.11 station and no additional hardware is required. The estimate of the $Slot\_Utilization$ has to be maintained and updated by each station, during the backoff interval that precedes a transmission attempt. The DCC mechanism requires that the $Slot\_Utilization$ definition satisfies the following two conditions:

- $Slot\_Utilization \in [0..1]$: the value $Zero$ should indicate that no slots observed in the Backoff interval resulted as busy, while the value One should indicate that every slot available for transmission resulted as busy;
- the $Slot\_Utilization$ values should be distributed in the [0..1] interval, proportionally to the network contention level (e.g. the rate of busy slots compared to the total number of observed slots available for transmission).

The 802.11 Standard mechanism assumes every station able to perform a Carrier Sensing activity such that the status (idle or busy) of the channel is known to each station, see Fig. 2. This makes possible to obtain a $Slot\_Utilization$ definition that satisfies the DCC requirements. It is worth noting that the $Slot\_Utilization$ estimate does not require to distinguish between successful transmissions and collisions. Each whole transmission (or collision)

[3] Collisions could occur even with few stations, so the contention indication obtained could be overestimated.

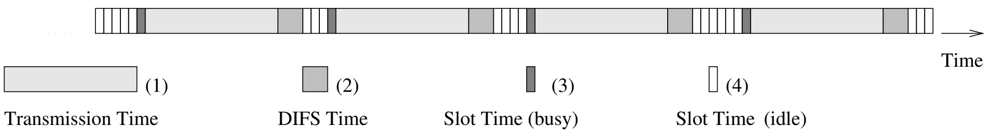

FIG. 2. A sample of the channel state observed by each station.

sensed on the channel assumes the role of a busy-slot indication. Specifically, during its own Backoff interval, every station counts the number of transmission attempts it observes on the channel (Num_Busy_Slots), and then divides this number by the total number of slots available for transmission observed on the channel ($Num\_Available\_Slots$). Hence, a simple and intuitive definition of the $Slot\_Utilization$ estimate is given by:

$$Slot\_Utilization = \frac{Num\_Busy\_Slots}{Num\_Available\_Slots} \quad . \tag{1}$$

It is interesting to observe that the $Slot\_Utilization$ (1) provides a lower bound for the actual contention level of the channel. In fact, as some stations may transmit in the same slot, it provides a lower bound to the real number of stations that have tried to access the channel during the last observed Backoff interval. If the value of the $Slot\_Utilization$ is high (i.e. close to One), this implies that during the last backoff interval we were in presence of a high contention level. By assuming a time locality of contention, such an information provides an indication about the state of the channel in successive slots. In general, we cannot say the opposite, i.e. if the value of the $Slot\_Utilization$ is low, then we were dealing with a low level of congestion, because of stochastic variability in slot selection[4]. Anyway, a low $Slot\_Utilization$ generally indicates a low contention level. The reactions that the proposed mechanism for congestion control performs, are based on the above considerations. The $Slot\_Utilization$ estimate is an useful indicator of the state of the channel and it guarantees that the DCC mechanism activates only when a real congested condition occurs. The stochastic variability of the $Slot\_Utilization$ estimate can only attenuate the mechanism's effects, and never amplify them in scenarios where no congestion occurs. The $Slot\_Utilization$ estimate constitutes the measure of the channel's congestion level that every station in the system will utilize to determine its behavior.

To summarize, the positive aspects of the $Slot\_Utilization$ estimate are: it constitutes a contention indication before a station performs a transmission (without the need to experience collisions), it is completely distributed, it provides a meaningful indication of the actual contention level, and it is obtained with no cost. Furthermore, the $Slot\_Utilization$ estimate can be directly obtained from the Standard, because all the information required is already available to each station.

The use of the $Slot\_Utilization$ estimate (1) presents a problem in the 802.11 Standard, due to the possible zero assignment for the $Backoff\_Counter$. A station which obtains a zero-backoff value is authorized to perform its transmission attempt immediately after the first idle DIFS interval it senses on the channel. This implies that such a station will perform an access to the channel without obtaining any information about the congestion

[4] As said before, the $Slot\_Utilization$ (1) is only a lower bound of the contention level, in fact a busy slot observed could have been selected for transmission by an arbitrary number of colliding stations.

level of the channel. Moreover, such stations could cause the starvation of remaining stations, because their transmissions (immediately after a DIFS) would cause no idle slots to be seen on the channel, and hence the remaining stations cannot decrement their backoff counters. A possible solution of this problem is based on the conservative assumption that the default $Slot\_Utilization$ level is One.

3.2. Management of contention in DCC

In this section, we present the use of the $Slot\_Utilization$ estimate (1) in a distributed contention-control mechanism, named DCC. To this end, we introduce a parameter, local to each station, which constitutes the core of the proposed mechanism. This parameter, named *Probability of Transmission* (P_T), allows to realize a filter between the standard access scheme and the physical layer, as shown in Fig. 3. This filter represents the opportunity

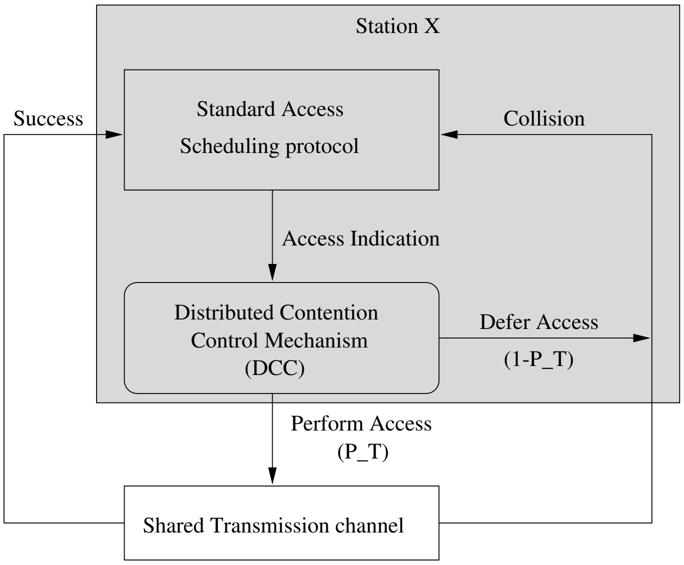

FIG. 3. Architectural collocation of DCC

to perform a transmission attempt. It is evaluated by a station to decide to perform or to defer the scheduled transmission attempt. This decision is based on the information gathered during the backoff interval. In few words, the only reasonable behavior of a station which knows that there are few possibilities for a successful transmission, is to defer its transmission attempt[5]. For this reason, according to DCC, each station should filter its transmission attempt depending on the value of the *Probability of Transmission* parameter, that is dependent on the contention level of the channel. Specifically, after the $Slot\_Utilization$ estimation (1), each station computes the *Probability of Transmission* value, and adopts it to evaluate the opportunity to access the shared channel. If the station decides to defer the transmission, it reschedules a new attempt, in the same way it does when a collision occurs (Fig. 3). Hence, this mechanism realizes a contention reduction, on a statistical basis, and reduces the collision probability with respect to the Standard protocol.

[5] This would also contribute to Power Saving, as in this way we save the power required by a transmission attempt which has a low probability to be successful.

Let us now analyze the expression adopted in the proposed mechanism to evaluate the *Probability of Transmission* (P_T). To control a congested system, it results opportune to reduce the number of accesses, i.e. the probability of collision. This can be easily obtained with a definition of the *Probability of Transmission* (P_T) which depends on the estimate of the $Slot\_Utilization$ (1) as:

$$P\_T(Slot\_Utilization) = (1 - Slot\_Utilization) \quad .$$

By definition, the $Slot\_Utilization$ (1) assumes values in the interval [0..1].
The $Slot\_Utilization$ represents the channel contention level, and its value is adopted in the DCC mechanism as a generalized inhibition on the accesses. It is worth noting that such a flat definition of the P_T would lead the system $Slot\_Utilization$ to fluctuate between two values, respectively $Zero$ and One, i.e. with no utilization of the channel. If the $Slot\_Utilization$ is high, then every user obtains a low P_T value, and this will reduce the future $Slot\_Utilization$. A low $Slot\_Utilization$ will cause every user to obtain a high P_T value, and after this a high $Slot\_Utilization$ will again appear on the channel.

To avoid this harmful fluctuating behavior, the P_T definition is extended by introducing a local parameter. The set of active stations is dynamically partitioned and each subset is associated with a different level of privilege for the accesses. For a tagged station, the privilege level for the transmission of a tagged frame is represented by the number of attempts already performed for that frame (Num_Att), including the current one. The Num_Att parameter is used in the DCC mechanism as a dynamic indicator of the privilege level achieved by a station:

$$P\_T(Slot\_Utilization, Num\_Att) = (1 - Slot\_Utilization^{Num\_Att}) \quad . \tag{2}$$

The lowest privilege is given to stations performing, for a given frame, the first transmission attempt, while the privilege level linearly increases with the number of collisions experienced. The aim is to privilege old transmission requests, obtaining a queue-emptying behavior for the system. The probability to perform an access should obviously increase with the increase of the user's waiting time. To this end, the effect of the Num_Att parameter represents a reduction of the inhibition power of the mechanism, for users with high Num_Att values. It is worth noting that the DCC's effect is probabilistic and not threshold-based. This fact allows a smoothed effect without introducing blocking behavior and reducing the risk of fluctuations. To better understand what we obtain, we can observe the Fig. 4. In Fig. 4 we show the P_T curves (2) as a function of the $Slot\_Utilization$ estimate, for different values of the number of attempts performed, Num_Att. Assuming a $Slot\_Utilization$ close to $Zero$, e.g. 0.2, we can observe that each station obtains a P_T close to One, independently by its Num_Att value. This means that the DCC mechanism has no effect on the system, and each user performs its accesses following the standard indications, without any additional contention control. This point is significant, because it means that no overhead is introduced by DCC in low-load scenarios. The DCC mechanism becomes effective when the $Slot\_Utilization$ grows. For example, by assuming a $Slot\_Utilization$ close to One, e.g. 0.8, we see that stations with the highest Num_Att value obtain a P_T value greater than the other stations. The priority order (connected to the local parameter Num_Att) preserves when it is mapped on the privilege order connected to the P_T parameter. It is worth noting that the Num_Att mechanism is completely

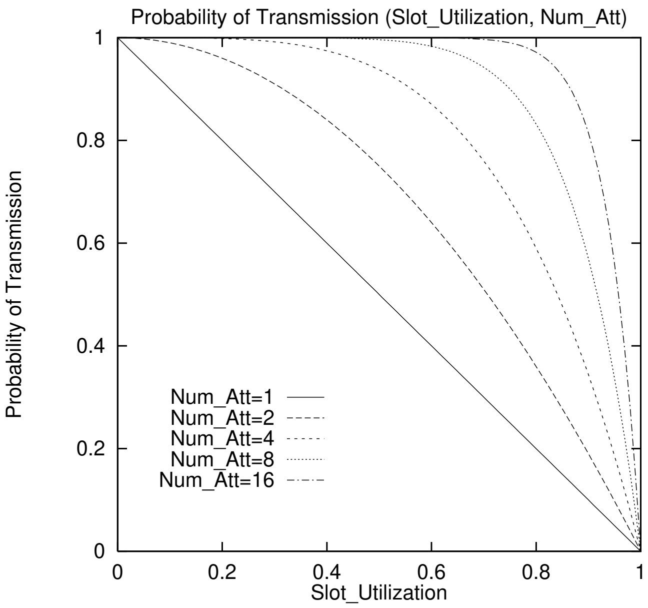

FIG. 4. Probability of Transmission

distributed, and each station does not have to worry about the other Num_Att values. The $Slot\_Utilization$ of the system, under this mechanism, never reaches the value One. In this case the inhibition mechanism would perform asymptotically, by reducing the P_T for each user, see Fig. 4. By opportunely managing the definition (2) of the *Probability of Transmission* (P_T), we could force any fixed limit as the asymptotic limit for the contention level. Hence, the DCC mechanism can be tuned to guarantee that the network operates within any fixed contention limit. This aspect is currently under investigation [2, 3].

When the system behavior is controlled by the *Probability of Transmission* values, the stations with the highest privilege have the fastest reactions when the $Slot\_Utilization$ decrements. This means a capture of the channel by the stations with the highest access privilege level. The low-privilege stations consider the deferring of their transmission attempt as a collision, and behave accordingly to the Standard (Fig. 3). Thus, when a station defers its transmission, it increases its privilege level (by increasing Num_Att), and hence it maintains a privilege with respect to younger (from the scheduling standpoint) transmission attempts.

To implement the DCC mechanism on top of the Basic Access mechanism of the Standard 802.11 DCF, we simply need to insert a new condition (i.e. the filter) before the access indication reaches the hardware (Fig. 3). Specifically, the DCC algorithm adopted by each station is sketched in Table 1.

To sum up, we realized a statistical based, distributed access-authorization mechanism devoted to obtain a contention reduction, when it is needed. Our mechanism is designed to lead to a capture of the resource by those users with the highest access delay among those waiting for transmission. The mechanism is also adaptive to the channel congestion level, due to the feedback effect connected to the $Slot\_Utilization$ estimate. Finally, it

Table 1: a sketch of the DCC algorithm.

```
...if (Backoff_Counter = 0)   /* The Slot for the transmission is reached */
        then
            calculate the Slot_Utilization estimate
            calculate the Probability of Transmission P_T(Slot_Utilization, Num_Att)
            if (RND() < P_T(Slot_Utilization, Num_Att))
                    then
                            BY-PASS the transmission indication to the HW
                    else
                            DEFER the transmission
            if (transmission deferred) or (Collision occurred)
                    then
                            NOTIFY(collision occurred)     /* schedule a new retransmission */
```

realizes an automatic tuning of the system behavior, without overheads, and with good characteristics from the performance standpoint, as we show in next sections.

The station privilege level depends in our mechanism on the value of the Num_Att parameter. A good management of such a parameter is thus necessary. To this end it is important to accurately define how a station's privilege increases. Simulation studies have shown that a fast increment of Num_Att makes flat the priority effect. To maintain the effectiveness of the privilege mechanism, it is useful to delay the effect of the Num_Att parameter up until the CW_Size reaches the CW_Size_MAX value. After this point, the Num_Att parameter is incremented by One only after each true transmission attempt. Specifically, at the beginning of a frame transmission, every time a transmission is deferred, it produces the increment of the CW_Size parameter. Once the CW_Size reached the maximum value, and the transmission is still to be performed, each unsuccessful transmission attempt implies the increment of the privilege level connected to Num_Att. No increment of Num_Att is done when a station decides by itself to defer a transmission attempt, according to the *Probability of Transmission* value.

4. BEHAVIOR OF THE SYSTEM

In this section, we analyze the behavior of the DCC mechanism in a network in which a set of M stations operates in asymptotic conditions (i.e. with continuous transmission requirements). Our study is based on some basic principles of the *Equilibrium Point Analysis (EPA)* [18]. The study focuses on the observation of the channel $Slot\_Utilization$ level, because in our mechanism it is the main variable that drives the system behavior. The analysis is based on the analytical expression of the future channel $Slot\_Utilization$ level (S_U_{next}) given the current $Slot\_Utilization$ level (S_U_{curr}), and the system's offered load. We subdivide the stations according to the following equivalence relationship ($\equiv$):

$$\text{stationA} \equiv \text{stationB} \Leftrightarrow \text{stationA.Num\_Att} = \text{stationB.Num\_Att} \qquad .$$

This partition gives the distribution of stations among classes related with their transmission rights. In the following we denote with H the number of equivalence classes. Assuming that, at a given time, each station has an estimate of the current value of the channel contention S_U_{curr}, we can compute the next future $S\_U_{next}(S\_U_{curr})$ contention level on the channel. This is done by calculating the probability of "at least one transmission" in the next slot on the channel.

We define a vector $\vec{h}$ such that h_i, $i \in [1..H]$, represents the number of stations trying to perform their i-th transmission attempt at the given time, and $\sum_{i=1}^{H} h_i = M$. We also define a vector $\vec{cw}$, such that cw_i, $i \in [1..H]$, maintains the Contention Window size of a station belonging to class i, i.e.

$$cw_i = min(CW\_Size\_MAX, CW\_Size\_min * 2^{i-1}) \quad .$$

Given Num_Att and S_U_{curr}, the probability for a given station to access the next slot, say $P_{acc\_next}(S\_U_{curr}, Num\_Att)$, is:

$$P_{acc\_next}(S\_U_{curr}, Num\_Att) = P\_T(S\_U_{curr}, Num\_Att) * P_{sel}(Num\_Att)$$

where

- $P_{sel}(Num\_Att)$ is approximated with the probability to select a given slot among those belonging to the station's Contention Window, i.e. $P_{sel}(Num\_Att) = 1/cw_{Num\_Att}$, and
- $P\_T(S\_U_{curr}, Num\_Att) = (1 - S\_U_{curr}{}^{Num\_Att})$ is the *Probability of Transmission* (2) in the scheduled slot.

In this way, given $(1 - P_{acc\_next}(S\_U_{curr}, i))^{h_i}$ the probability that "no class-i stations transmit" in the next slot, we obtain the probability of "at least one transmission attempt" in the next slot interval, i.e. the $S\_U_{next}(S\_U_{curr})$ by the following expression:

$$S\_U_{next}(S\_U_{curr}) = (1 - \prod_{i=1}^{H} (1 - P_{acc\_next}(S\_U_{curr}, i))^{h_i}) \quad . \tag{3}$$

Now, let us assume that one or more equilibrium points exist for the system behavior, from the *Slot Utilization* point of view. To study these equilibrium points, we impose that the current level of *Slot Utilization* be equal to the future level, i.e.

$$S\_U_{next}(S\_U_{curr}) = S\_U_{curr} \quad . \tag{4}$$

The number of S_U_{curr} values satisfying the above condition (4), identifies the number of possible equilibrium points. These equilibrium points are shown in Fig. 5, and they correspond to the intersections between the line representing the system's trend of the *Slot Utilization* level $S\_U_{next}(S\_U_{curr})$ and the load line representing the current *Slot Utilization* level S_U_{curr} on the channel.

The Fig. 5 shows a set of lines which represent the values of $S\_U_{next}(S\_U_{curr})$ (3) for various values of M. To plot the curves we assumed that $H = 10$ and that the M stations are equally distributed over the H classes. The points on the straight line, such that $S\_U_{next}() = S\_U_{curr}$, represent the equilibrium condition. The other $S\_U_{next}(S\_U_{curr})$ points should be interpreted as follows:

- if the value of $S\_U_{next}(S\_U_{curr})$ is over the corresponding value of S_U_{curr} (i.e. over the straight line), then the expected *Slot Utilization* in the near future will grow, and the system will be pushed to work towards the S_U_{curr} value corresponding to the point of intersection between the lines.

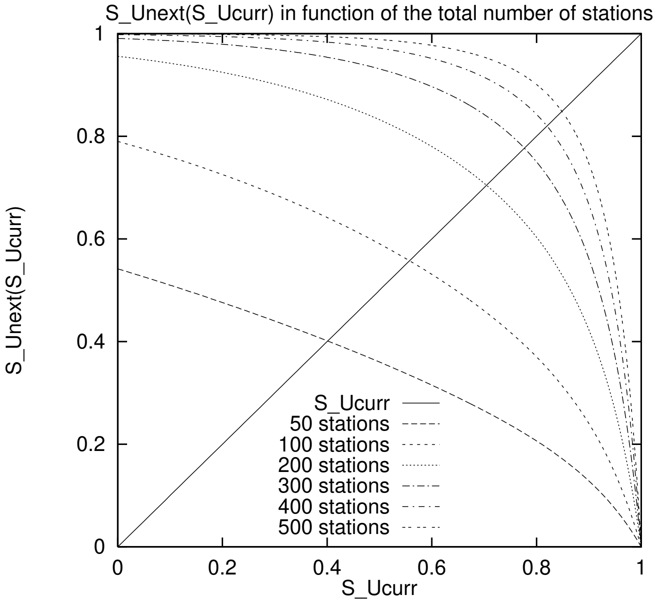

FIG. 5. System's behavior

• if the value of $S\_U_{next}(S\_U_{curr})$ is under the corresponding value of S_U_{curr} (i.e. under the straight line), then the expected *Slot Utilization* in the near future will decrease, and again the system will be pushed to work towards the S_U_{curr} value corresponding to the point of intersection between the lines.

We can observe that, for the given system configuration, there is only one convergence point. This fact can be proved by demonstrating that, for each system considered, the $S\_U_{next}(S\_U_{curr})$ expression (3) is monotonic decreasing, so there is a single point of intersection with the straight line representing the equilibrium condition (4). The intersection point is an indication of the level of *Slot Utilization* that the system would produce in the near future. These points give an indication about the congestion reaction and control of the system. They assure that the system behavior has no convergence points corresponding to a very low channel utilization. This analysis also demonstrates that, under the effect of the proposed DCC mechanism, the *Slot Utilization* convergence-point reaches the value One only when M grows to infinity. Such a result indicates that the proposed DCC mechanism introduces the possibility to avoid a network collapse. Moreover, by the opportune definition of the *Probability of Transmission* expression (2), it would be possible to define a general asymptotic contention limit for the *Slot Utilization* generated on the channel. This could be used to obtain an arbitrary reduction of the risk of a network collapse.

5. SIMULATION RESULTS

In this section we present a set of results, obtained via discrete event simulation, to compare the performance of the Standard 802.11 DCF with the one of the Standard protocol enhanced with the DCC mechanism. To perform each simulation, it has been developed a simulation model with the Research Queueing Package Version 3 (RESQ) simulation tool [14]. The physical parameters of the system have been taken by literature [1, 4, 7], and are specified in Table 2. Active stations are assumed to have continuous transmission

Table 2: Simulation parameters for the considered system

NUMBER OF STATIONS	(VARIABLE) 4 . . 512	CW_Size_min	8 Slots
Channel Rate	2 Mb/s	CW_Size_MAX	256 Slots
Frame Payload	(CONSTANT) 1024 B	Ack size	50 B
Header size	34 B	Slot Time	20 µSec
S I F S	20 µSec	D I F S	50 µSec

requirements. In our simulation experiments the number of active stations changes from 4 to 512, to represent different levels of congestion and thus to investigate the scalability of the proposed mechanism. Each time an active station performs a successful transmission, it resets the whole mechanism and immediately reschedules a new frame transmission. In our work we don't consider the *Hidden Terminal* problem [5, 6, 8, 20], and we assume that the only cause of collision between two or more transmitting stations is the selection of the same slot for their transmissions. According to the above assumptions, we focus our attention to study the MAC protocol behavior from the congestion-control and congestion-reaction points of view. Results presented in this work have been obtained by assuming a constant frame size. Ongoing studies with a variable frame size confirm and re-inforce the results presented in this paper. For all simulation results we also reported their confidence intervals corresponding to a 95% confidence level.

5.1. Number of collisions for each successful transmission

The first performance index we evaluated is the mean number of collisions experienced by a station before a successful transmission. This observation gives a meaningful indication of the reduction obtained in the probability of collision. As shown in Fig. 6, the use of the DCC mechanism allows a significant reduction in the number of collisions[6]. This reduction is more marked as the contention level increases. Moreover, it has to be noted that the congestion reaction of the proposed mechanism guarantees a linear increase of the number of collisions with the increase in the number of active stations, even when the number of stations in the system exceeds the reaction capability of the Binary Exponential Backoff mechanism, i.e. the number of stations exceeds CW_Size_MAX. On the contrary, when the number of stations becomes greater than CW_Size_MAX (256), it is possible to observe a knee in the number-of-collisions curve corresponding to the Standard.

5.2. Mean Access Delay

Fig. 7 shows the Mean Access Delay in a system with different levels of congestion. It is simple to verify that the proposed mechanism significantly improves this performance index, with respect to the Standard protocol.

It is worth noting that no overhead is introduced, (i.e. the mean access delay with DCC is never higher than the standard-protocol access delay) even in scenarios with few stations (Fig. 8). This fact is important, because it confirms our expectation that it is possible to reduce the Mean Access Delay by introducing ad-hoc delays in the transmission attempts, when the contention level is high. Even if this could appear as a contradiction it results useful, because of the combined effect of *i)* the reduction of the contention level obtained

[6] It is worth noting that the reduction in the number of collisions implies a corresponding reduction in the power consumption required for the transmission of a frame [2].

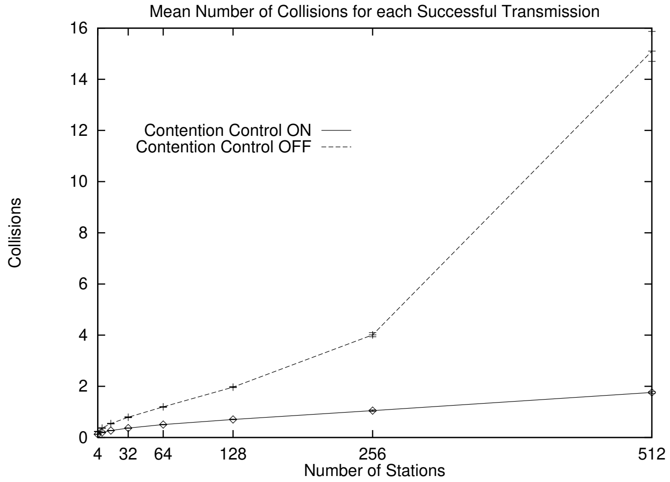

FIG. 6. Mean number of collisions for each successful transmission.

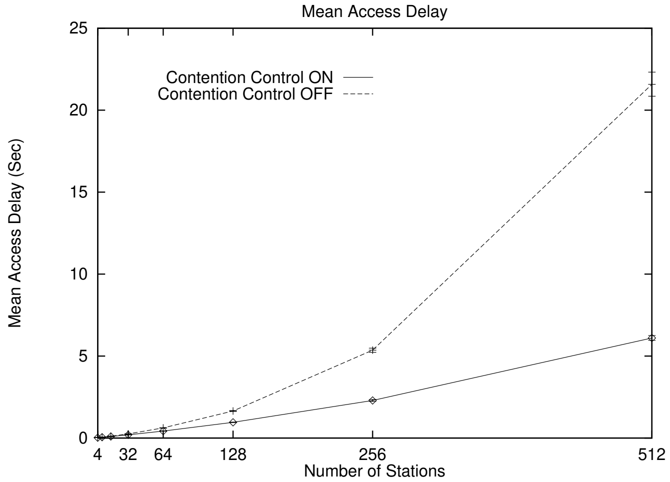

FIG. 7. Mean Access Delay

by introducing such delays, and *ii)* the priority mechanism which increases the privilege in accessing the system as the station access-delay increases.

5.3. 99-th percentile of Access Delay

A further meaningful parameter to be investigated is the 99-th percentile of Access Delay (Fig. 9). This parameter confirms the reduction in the access delay, assuring that the priority mechanism connected to the number of attempts performed results effective. The study of this index allows us to verify that the reduction in the Mean Access Delay was not obtained at the expense of increasing the delay in a subset of unlucky stations. This is also a good indication about the reduction in the starvation risk achieved with the DCC mechanism.

5.4. Channel utilization

An important indication about the effectiveness of the proposed mechanism is obtained by analyzing the channel utilization level (Fig. 10). It has to be noted the great improvement

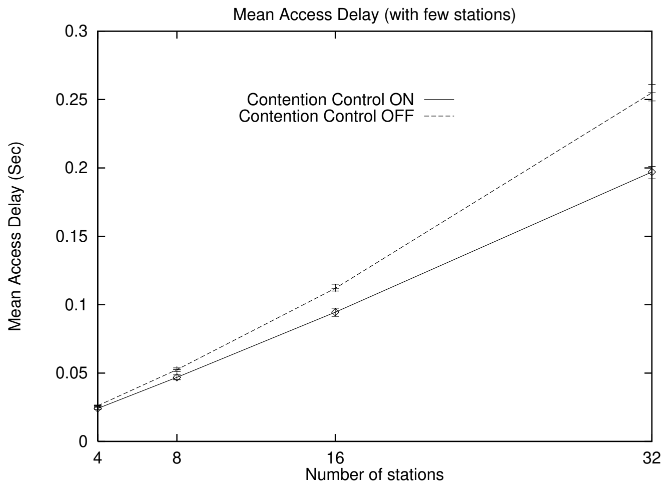

FIG. 8. Mean Access Delay (zoomed)

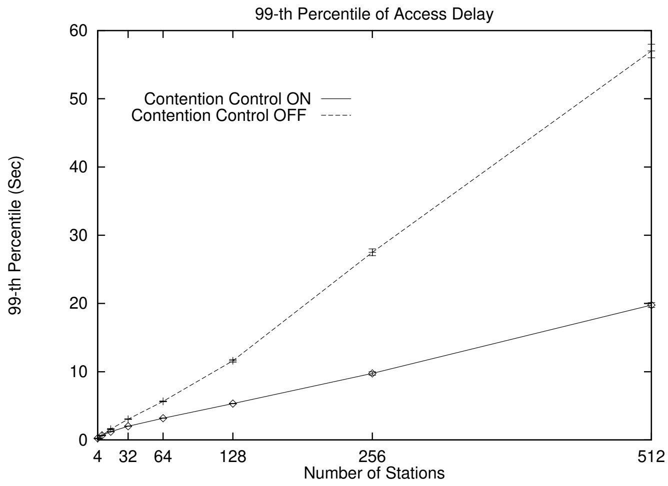

FIG. 9. 99-th percentile of Access Delay

obtained by adopting DCC in congested systems, and the absence of overheads (intended as reduction of the available channel bandwidth) introduced by DCC when there are few active stations.

5.5. DCC behavior with bursty arrivals

Fig. 11 and 12 show the analysis of the system reaction when the congestion level drastically changes. Specifically, in Fig. 11, we considered a system with 128 steady-state active stations.

At a given time instant (i.e. after about 6 seconds in Fig. 11) 128 additional stations become active, each one with a single frame to transmit. In our experiments we measured the time the system takes to satisfy all the bursty traffic represented by the 128 additional transmission requirements. In Fig. 12 we report the results related to an initially empty system in which a burst of 256 transmission requirements suddenly appears.

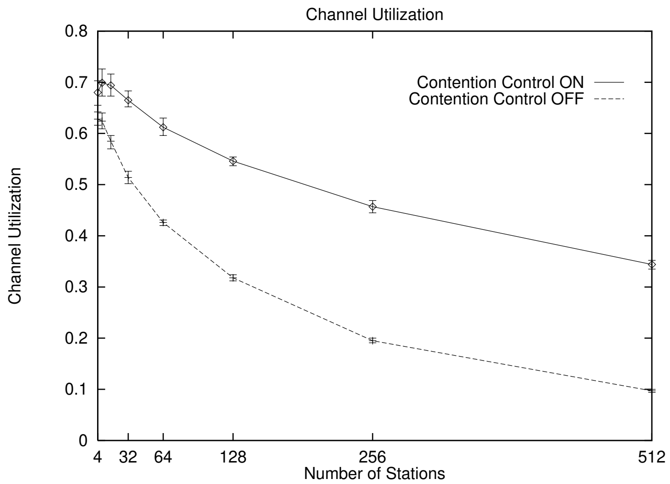

FIG. 10. Channel Utilization

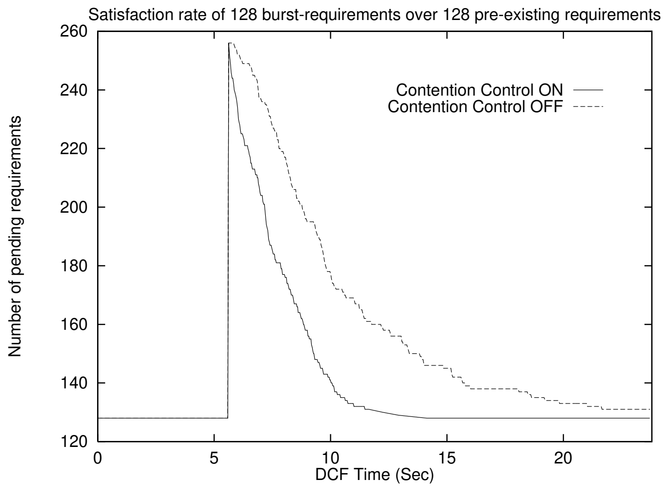

FIG. 11. 128 bursty arrivals (congested system)

In both scenarios, the system which adopts the DCC mechanism shows a better and faster congestion reaction, with respect to the standard system. These results point out the DCC capability to contrast the congestion growth, and to provide a fast congestion reaction. Note that the DCC higher speed in satisfying the transmission requirements is mainly due to the higher channel utilization obtained, with respect ot the Standard. Note also the almost complete absence of tail in the DCC mechanism's curve (Fig. 11), which demonstrates the privilege-mechanism effect (based on the Num_Att parameter), that produces a queue-emptying behavior.

6. PRIORITY MECHANISM

It is interesting to note that it is possible to extend the basic DCC mechanism with a priority mechanism. The priority mechanism introduces several classes of traffic in the network, without requiring any negotiation. Each station could select the priority level for a given frame, e.g. depending on the application requirements. This could be thought as a

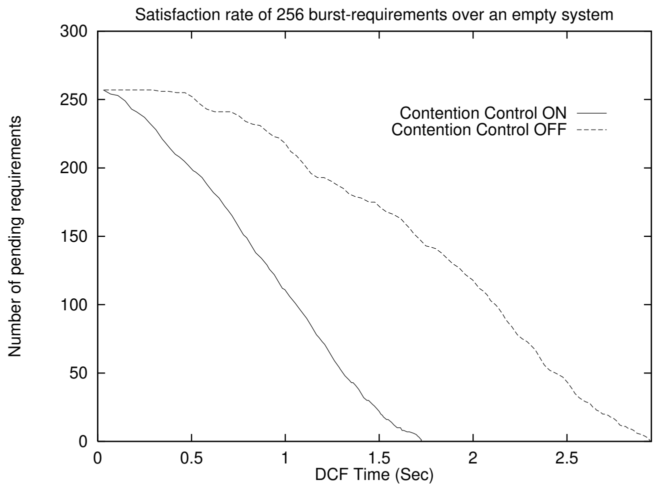

FIG. 12. 256 bursty arrivals (empty system)

way to map on the wireless MAC protocol the requirements of the Type of Service (ToS) bits of the IPv4 protocol [17]. It is worth noting that, if the system is not congested, a frame can be typically delivered in a limited time. The priority mechanism becomes effective in congested scenarios, allowing high-priority frames to be transmitted with an enhanced QoS. Specifically, this mechanism forces low-priority traffic to reduce its throughput by thus increasing the portion of bandwidth available for the high-priority traffic.

The extension of DCC with priorities could be obtained with a little change in the *Probability of Transmission* definition (2), simply adding a specific local parameter, named *Priority Level* $(Prior\_Lev)$, with values greater or equal to 1 (default = 1):

$$\begin{array}{l} P\_T(Slot\_Utilization, Num\_Att, Prior\_Lev) = \\ \qquad (1 - Slot\_Utilization^{Prior\_Lev * Num\_Att}) \end{array} \quad (5)$$

For each given frame to transmit, the $Prior\_Lev$ parameter could be mapped on the ToS values defined by the application level. When the $Prior\_Lev$ parameter is greater than 1, it introduces a fast increment in the priority of the station, with respect to the number of transmission attempts performed. Simulation studies of such a mechanism, presented in the next section, have shown its effectiveness. Simulations also show that some stations can improve the QoS they experience in congested systems, simply selecting a $Prior\_Lev$ value greater than One.

6.1. Performance Evaluation of the Priority mechanism

The Fig. 13 shows the effectiveness of the priority mechanism, in terms of the Mean Access Delay obtained by three sets of stations with traffic belonging to three different, self-determined, priority levels $(Prior\_Lev = 1, 2, 3)$.

The Fig. 14 confirms the effectiveness of the priority mechanism by comparing the 99-th percentile of the access delay, experienced by a small set of high-priority stations, to the 99-th percentile of the access delay experienced by the remaining stations. Specifically, the system considered is constituted by a set of 4 stations with $Prior\_Lev = 2$, 4 stations with

$Prior\_Lev = 3$, and a variable set of 10 up to 200 stations with no privileges ($Prior\_Lev$ = 1).

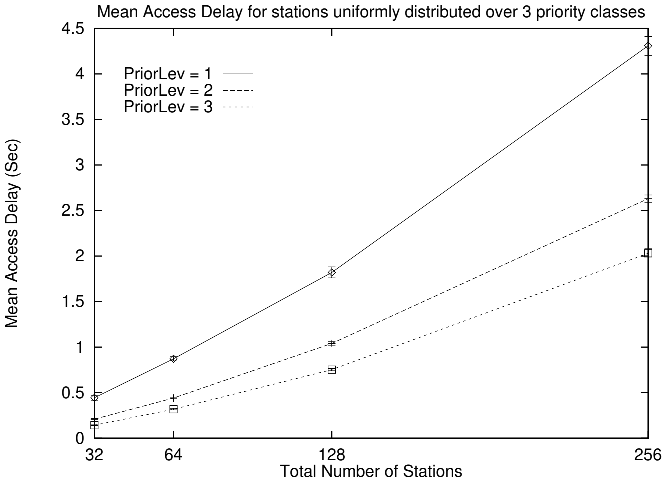

FIG. 13. Mean access delay with Priority levels

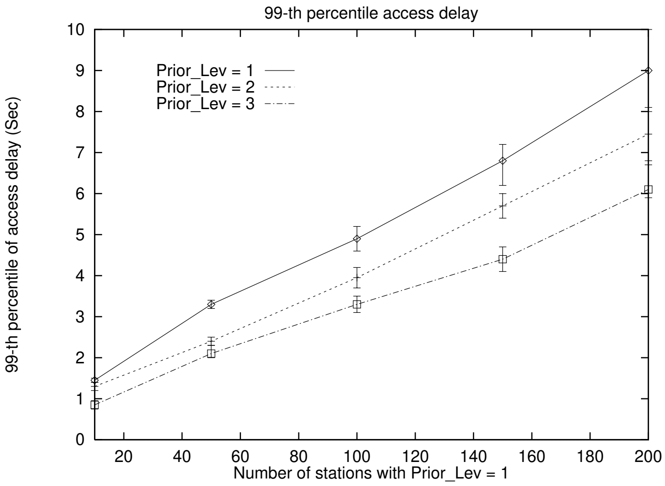

FIG. 14. 99-th percentile of access delay with Priority levels

Fig. 13 and 14 demonstrate that the priority level influences the access delay of all stations. A high-priority station can achieve a better access delay with respect to low-priority stations.

Fig. 15 and 16 show the throughput of each station depending on its *Priority Level*. Specifically, the throughput index is defined as follows: thr_k is the throughput of a given station with $Prior\_Lev = k$, normalized with respect to the total throughput of the system:

$$thr_k = \frac{throughput\_of\_given\_station\_with\_Prior\_Lev = k}{throughput\_of\_the\_system} \quad .$$

Let thr_{fair} be the normalized throughput in a fair system, with no priorities:

$$thr_{fair} = \frac{throughput\_of\_the\_system}{total\_number\_of\_stations} * \frac{1}{throughput\_of\_the\_system} = \frac{1}{total\_number\_of\_stations} \quad .$$

The throughput index we consider in the Fig. 15 and 16 is defined as:

$$T_k = thr_k / thr_{fair} \quad . \tag{6}$$

This index shows the enhancement in the QoS experienced by a station, by increasing its priority level. It is worth noting that a value One of this index (6) indicates a fair normalized throughput, i.e. the throughput achieved by each station in a fair access scheme with no priorities (e.g. a round-robin scheduling). A value greater than One indicates the enhancement of the QoS achieved by a station. For example, a throughput index (6) equal to 2 indicates that the corresponding station can achieve twice the throughput that it would obtain in a fair system with no priorities. On the other hand, index values less than One imply that the given station obtains a throughput lower than a fair throughput level.

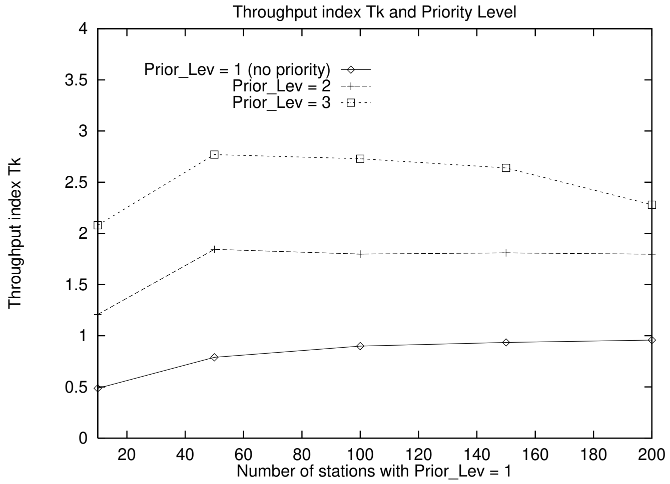

FIG. 15. Throughput index and Priority levels: system A

Fig. 15 and 16 show the impact of the *Priority Level* on the throughput achieved by a station. The reported values are obtained by averaging the indexes experienced by the stations belonging to the same class of priority. We considered two different systems in our simulations. The system A (Fig. 15) has a set of 4 stations with $Prior\_Lev = 2$, 4 stations with $Prior\_Lev = 3$, and a variable set of 10 up to 200 stations with $Prior\_Lev = 1$ (default). The system B (Fig. 16) has a set of 4 stations with $Prior\_Lev = 5$, 4 stations with $Prior\_Lev = 10$, and a variable set of 10 up to 200 stations with $Prior\_Lev = 1$ (default).

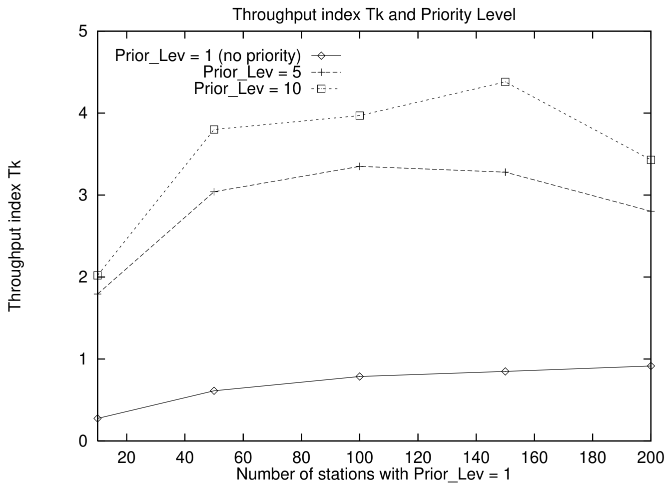

FIG. 16. Throughput index and Priority levels: system B

It can be seen that the priority mechanism results effective, by providing an increase in the station throughput, by increasing its *Priority Level*. Note that the additional throughput obtained by (few) priority stations marginally influences the throughput of (much more) low-priority stations. Even if it does not appear in the figures, simulation results show that the mechanism guarantees a fair access to the channel among all the stations belonging to the same priority class. As expected, by comparing the curves presented in Fig. 15 and 16, it results that increasing the value of the $Prior\_Lev$ parameter produces a throughput gain for a class of users, but additional studies are required to define the relationship between the priority level value and the throughput achieved by each class of traffic.

7. CONCLUSIONS AND FUTURE RESEARCHES

In this paper we proposed and extensively evaluated a distributed mechanism, named DCC, for the contention reduction in random-access protocols. Specifically, we have shown the effectiveness of this mechanism in improving the performance of the IEEE 802.11 protocol. The proposed DCC mechanism results very effective in terms of contention control, congestion avoidance and congestion recovery, and it is completely distributed. It is able to identify in advance (i.e. without generating collisions) a congestion condition and it behaves in a fully adaptive way, leading to a system's behavior characterized by high stability, and with no overheads introduced in low-contention conditions. The analysis of the system under the proposed DCC mechanism shows the absence of harmful fluctuations, both during and after the transient phases. No further hardware, communication, power consumption or significant time is required with respect to the Standard protocol. The DCC mechanism can be used on top the 802.11 Standard DCF, with no significant changes required. Among the advantages of the proposed mechanism, we can observe an improvement in the channel utilization, and a decrease of the mean and 99-th percentile of the access delay. Furthermore, our mechanism guarantees a better response of the system to transient changes in the contention level, and it reduces the risk of starvation. These advantages increase with the increase in the system's congestion level. It also results possible to introduce a distributed priority mechanism, simple and effective, which allows each station to self-determine its priority level.

The impact of Hidden Terminals has not been considered in the analysis of the proposed mechanism. However, we can observe that the DCC mechanism requires only a rough investigation of the *Slot Utilization* level of the system, and the presence of Hidden Terminals can introduce a biasing of this estimate. For example, if all stations are active, and if for each station a 10% of terminals are hidden, the impact on the $Slot\_Utilization$ estimate could result in a biased $Slot\_Utilization$ whose value would be lower than the real one up to 10%. This would marginally reduce the effectiveness of the DCC mechanism in congested systems. Note also that the extreme worst case (i.e. 100% hidden terminals) would cause the $Slot\_Utilization$ estimate to be zero, and the *Probability of Transmission* (P_T) to be one. Hence, in the extreme case from the Hidden Terminal standpoint, the DCC mechanism performs like the Standard 802.11 access scheme.

We are now investigating the evolution of the DCC mechanism to capture the relationship between the traffic characteristics (mainly the average message length), the maximum channel utilization level [3], and the optimal power saving characteristics [2]. This evolution is based both on an analytical study of the MAC protocol capacity similar to the one proposed in [4], and on a low cost estimate of the size of the transmitted frames. Future research on DCC mechanism will include the adoption of smoothing functions to estimate the $Slot\_Utilization$, the design of different scheduling functions, and the exploitation of the priority level mechanism for RTS/CTS packets [8] in order to privilege their transmission.

It is worth noting that the effectiveness of DCC could be exploited in other contexts of distributed systems involving resource sharing, provided that they adopt some kind of random access control, and that it is possible to obtain some indications about the contention level in accessing such resources.

REFERENCES

1. Bianchi G., Fratta L. and Olivieri M. (1996). “Performance Evaluation and Enhancement of the CSMA/CA MAC protocol for 802.11 Wireless LANs”, Proc. of PIMRC 1996, Oct. 1996, Taipei, Taiwan, pp. 392-396.
2. Bononi L., Conti M. and Donatiello L. (1999). “A Distributed Contention Control Mechanism for Power Saving in random-access Ad-Hoc Wireless Local Area Networks”, Proc. of MOMUC’99, San Diego, CA, November 1999.
3. Bononi L., Conti M. and Gregori E. (2000). “Design and Performance Evaluation of an Asymptotically Optimal Backoff Algorithm for IEEE 802.11 Wireless LANs”, Proc. of HICCS-33, Maui, Hawaii, January 2000.
4. Cali’ F., Conti M. and Gregori E. (1998). “IEEE 802.11 Wireless LAN: Capacity Analysis and Protocol Enhancement”, Proc. INFOCOM’98, San Francisco, CA, March 29 - April 2, 1998, pp. 142-149.
5. Chen K.C. (1994). “Medium Access Control of Wireless LANs for Mobile Computing”, IEEE Networks, 9-10/1994.
6. Chhaya H.S. (1996). “Performance Evaluation of the IEEE 802.11 MAC Protocol for Wireless LANs”, Master Thesis, Graduate School of Illinois Institute of Technology, May 1996.
7. Crow B.P. (1996). “Performance Evaluation of the IEEE 802.11 Wireless Local Area Network Protocol”, Master thesis. University of Arizona, 1996.
8. IEEE 802.11 Std. (1997). IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Nov. 1997.
9. Gallagher R.G. (1986). “A perspective on multiaccess channels”, IEEE Trans. Information Theory, vol. IT-31, No.2, 3/1985, pp. 124-142.
10. Georgiadis L. and Papantoni-Kazakos P. (1985). “Limited Feedback Sensing Algorithms for the Packet Broadcast channel”, IEEE Trans. on Info. Theory, Vol IT-31, No.2, 3/1985, pp. 280-294.
11. Goodman J., Greenberg A.G., Madras N. and March P. (1985). “Stability of Binary Exponential Backoff”, Proc. of the 17-th Annual ACM Symp. on Theory of Comp., Providence, May 1985.
12. Hammond J.L. and O’Reilly P.J.P. (1988). *Performance Analysis of Local Computer Networks*, Addison-Wesley, 1988.

13. Hastad J., Leighton T. and Rogoff B. (1996). “Analysis of Backoff Protocols for Multiple Access Channels”, Siam J. Computing vol. 25, No. 4, 8/1996, pp. 740-774.

14. MacNair E.A. and Gordon R.F. (1991). “An Introduction to the RESearch Queueing Package for Modeling Computer Systems and Communication Networks”, IBM Research Report RA-208, Yorktown Heights, New York (June 1991).

15. Rivest R.L. (1987). “Network Control by Bayesian Broadcast”, IEEE Trans. on Info. Theory, Vol IT-33, No.3, 5/1987, pp. 323-328.

16. Stallings W. (1996). *Local & Metropolitan Area Networks*, Fifth Edition, Prentice Hall 1996, pp. 356-383.

17. Stevens W.R. (1994). *TCP/IP Illustrated*, Vol. 1, Addison Wesley, 1994.

18. Tasaka S. (1986). *Performance Analysis of Multiple Access Protocols*, MIT Press 1986.

19. Tsitsiklis J.N. (1987). “Analysis on a multiaccess Control Scheme”, IEEE Trans. on Auto. Contr., Vol AC-32, No. 11, 11/1987, pp. 1017-1020.

20. Weinmiller J., Woesner H., Ebert J.P. and Wolisz A. (1996). “Analyzing and tuning the Distributed Coordination Function in the IEEE 802.11 DFWMAC Draft Standard”, Proc. Int. Workshop on Modeling, MASCOT 96, San Jose, CA.