QoS provided by the IEEE 802.11 wireless LAN to advanced data applications: a simulation analysis

G. Anastasi and L. Lenzini

Department of Information Engineering, University of Pisa, Via Diotisalvi 2, 56126 Pisa, Italy

IEEE 802.11 is a Media Access Control (MAC) protocol which has been standardized by IEEE for Wireless Local Area Networks (WLANs). The IEEE 802.11 MAC protocol offers two types of services to its users: *synchronous* and *asynchronous*. This paper presents an in-depth analysis, by simulation, of the asynchronous part alone. The analysis is performed by considering station data traffic patterns (hereafter *advanced* data traffic) which have a very similar shape to traffic generated by WWW applications. We carried out the simulation by taking into consideration two classes of scenarios: balanced and unbalanced. In the former class each station has the same offered load while in the latter class a specific station is more loaded than the others. Our conclusion is that the IEEE 802.11 MAC protocol performs satisfactorily for both classes of scenarios, although performance measures with advanced traffic are worse than the corresponding performance measures with Poissonian traffic. Furthermore, we broadened our analysis to include higher medium capacities than those planned (i.e., 1 and 2 Mbit/sec) up to 10 Mbit/sec. This part of the analysis shows that the IEEE 802.11 MAC protocol is not adequate to work at speeds planned for the forthcoming ATM Wireless LAN.

1. Introduction

The IEEE 802.11 is a standard for a WLAN covering both physical and MAC layers [9,11]. A key issue of this standard is that a mobile host is able to communicate with any other mobile or wired host in a transparent manner. In other words, an IEEE 802.11 WLAN appears to layers above the MAC layer like any other IEEE 802.X LAN (e.g., Ethernet or Token Ring). This means, in particular, that the mobility aspects are handled at the MAC level or below.

A performance analysis for the IEEE 802.11 WLAN is reported in [2] where the authors take into account the decentralized nature of communication between stations, the possibility of capture, and the presence of hidden stations. They also study the impact of spatial characteristics on system performances. Both parts (synchronous and asynchronous) of the IEEE 802.11 MAC protocol have been analyzed in [13]. A comparison of the IEEE 802.11 MAC protocol with HIPERLAN [8], the ETSI proposal for wireless LANs, can be found in [14].

The aim of this paper is to further advance previous analysis on the asynchronous part of the IEEE 802.11 standard. In fact, since we expect relevant data applications serviced by WLANs to very much resemble today's WWW applications, in our performance analysis we use data traffic models which are very similar to those recently proposed in the literature [4,5] for these type of applications (i.e., WWW applications).

Furthermore, since discussions at an international level for a MAC protocol of an ATM Wireless LAN have just begun, we also analyze whether or not the IEEE 802.11 MAC protocol is suitable for operating at higher channel speeds.

Although data integrity is a key requirement for data transmission, we assume an error free radio channel despite the fact that we are dealing with a very unreliable environment. This is because in the present paper we are focusing on multiple access aspects of the IEEE 802.11 MAC protocol as measured by the average access delay experienced by packets transmitted over data connections.

The paper is organized as follows. Section 2 describes the IEEE 802.11 MAC protocol. Section 3 describes the simulation environment while the results are reported in sections 4 and 5. Section 6 concludes the paper.

2. IEEE 802.11 wireless network standard

In this section we only report those aspects of the standard which are relevant for our analysis. Specifically, section 2.1 introduces the terminology and topology adopted for an IEEE 802.11 WLAN, whereas section 2.2 describes the MAC protocol. See [9] for details.

2.1. Network topology

An IEEE 802.11 WLAN generally consists of *Basic Service Sets* (BSSs) which are interconnected by a *Distribution System* (DS) to form an Extended Service Set (ESS) as shown in figure 1.

Each BSS consists of a group of wireless stations which execute a *Distribution Function* (DF) to regulate the exclusive access to the shared wireless medium. Since the wireless medium is broadcast, each station can transmit directly to any other station in the same BSS. On the other hand, to transmit to stations belonging to a different BSS, stations pass through an *Access Point* (AP) which is an inter-working unit implementing both the IEEE 802.11 and the DS MAC protocols.

However, before a station can access the wireless medium it needs to be associated with an Access Point.

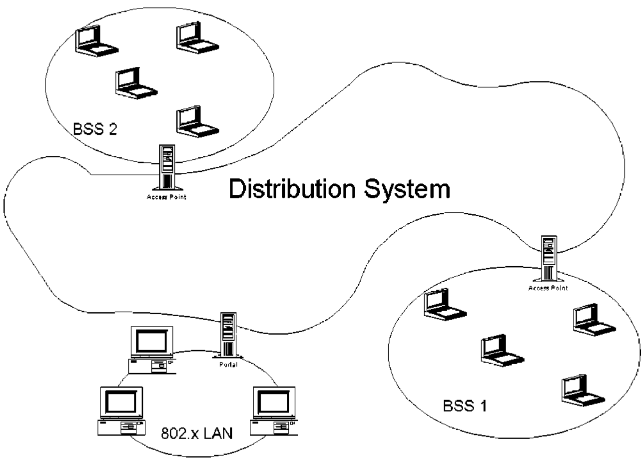

Figure 1. An IEEE 802.11 network topology.

A station can be associated with only one Access Point at a given time. The Distribution System supports mobility by providing the necessary services for handling destination mapping and the integration of BSS's in a manner that is transparent to stations. This means that hosts (either wireless or wired) do not need to know the physical location of other hosts for communications.

2.2. MAC protocol

The IEEE 802.11 MAC protocol provides two service types: *asynchronous* and *synchronous* (or, rather, *contention free*). These types of services can be provided on top of a variety of physical layers and for different data rates. The asynchronous type of service is always available whereas the contention free is optional.

The asynchronous type of service is provided by the *Distributed Coordination Function* (DCF) which implements the basic access method of the IEEE 802.11 MAC protocol and is also known as the *Carrier Sense Multiple Access with Collision Avoidance* (CSMA/CA) protocol.

The contention free service is provided by the *Point Coordination Function* (PCF) which basically implements a "polling" access method. The PCF uses a Point Coordinator, usually the Access Point, which cyclically polls stations giving them the opportunity to transmit. Unlike the DCF, the implementation of the PCF is not mandatory. Furthermore, the PCF itself relies on the asynchronous service provided by the DCF.

Since this paper does not analyze the contention free access method, in the following only the DCF will be described. Details about the PCF can be found in [9].

According to the DCF (see figure 2) a station must sense the medium before initiating the transmission of a packet. If the medium is sensed as being idle for a time interval greater than a *Distributed InterFrame Space* (DIFS) then the station transmits the packets. Otherwise, the transmission is deferred and the backoff process is started. Specifically, the station computes a random time interval, the *backoff interval*, uniformly distributed between zero and

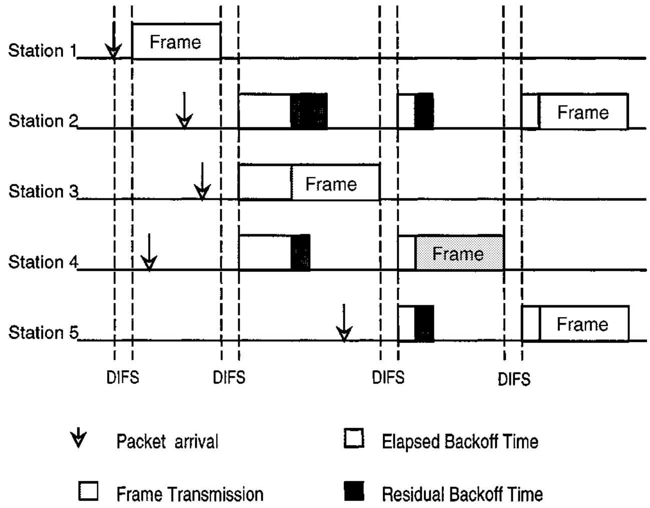

Figure 2. Basic access mechanism.

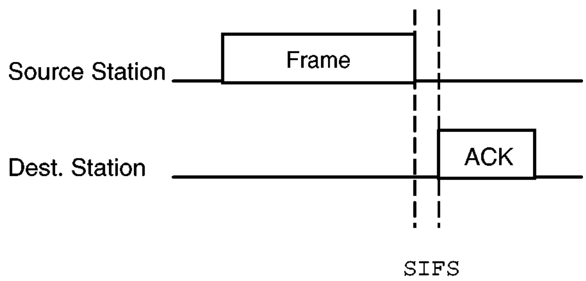

Figure 3. Acknowledgement mechanism.

a maximum called *Contention Window* (CW). This backoff interval is then used to initialize the *backoff timer*. This timer is decreased only when the medium is idle, whereas it is frozen when another station is transmitting. Specifically, each time the medium becomes idle, the station waits for a DIFS and then periodically decrements the backoff timer. The decrement period is referred to as the *slot-time* which corresponds to the maximum round-trip delay within the BSS and, hence, depends on the maximum BSS coverage.

As soon as the backoff timer expires, the station is authorized to access the medium. Obviously, a collision occurs if two or more stations start transmission simultaneously. Unlike wired networks (e.g., with CSMA/CD), in a wireless environment collision detection is not possible. Hence, as shown in figure 3, a positive acknowledgement is used to notify the sending station that the transmitted frame has been successfully received. The transmission of the acknowledgement is initiated at a time interval equal to the *Short InterFrame Space* (SIFS) after the end of the reception of the previous frame. Since the SIFS is, by definition, less than the DIFS[1] the receiving station does not need to sense the medium before transmitting the acknowledgement.

If the acknowledgement is not received the station assumes that the transmitted frame was not successfully received and, hence, schedules a retransmission and enters

[1] The DIFS is defined as DIFS = SIFS + 2 Slot-times.

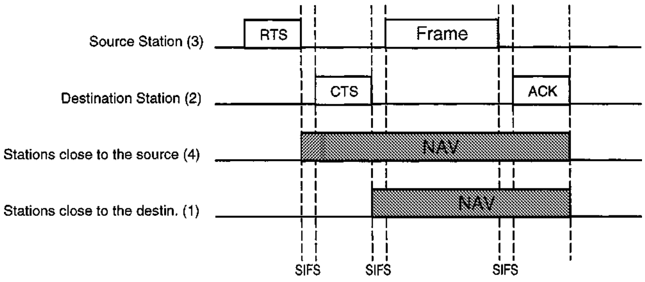

Figure 4. RTS/CTS mechanism.

the backoff process again. However, to reduce the probability of collisions, after each unsuccessful transmission attempt, the Contention Window is doubled until a predefined maximum (CWmax) is reached.

After a (successful or unsuccessful) frame transmission, if the station still has frames queued for transmission, it must execute a new backoff process.

In radio systems based on medium sensing, a phenomenon known as the *hidden station* problem may occur. This problem arises when a station is able to successfully receive frames from two different transmitters but the two transmitters cannot receive signals from each other. In this case a transmitter may sense the medium as being idle even if the other one is transmitting. This results in a collision at the receiving station.

To deal with the hidden station problem, the IEEE 802.11 MAC protocol includes an optional mechanism which is based on the exchange of two short control frames (see figure 4): a *Request To Send* (RTS) frame which is sent by a potential transmitter to the receiver and a *Clear To Send* (CTS) frame which is sent from the receiver in response to the received RTS frame. If the CTS frame is not received within a predefined time interval, the RTS frame is retransmitted by executing the backoff algorithm described above. After a successful exchange of the RTS and CTS frames, the data frame can be sent by the transmitter after waiting for a SIFS.

The RTS and CTS frames (as well as data and acknowledgement frames) include a *duration field* that specifies the time interval necessary to completely transmit the data frame and the related acknowledgement. This information is used by stations which can hear either the transmitter or the receiver to update their *Net Allocation Vector* (NAV), a timer which, unlike the backoff timer, is always decreased, i.e., irrespective of the medium status. Since stations which can hear either the transmitter or the receiver refrain from transmitting until their NAV has expired, the probability of a collision due to a hidden station occurring is reduced. Of course, the drawback of using the RTS/CTS mechanism is an increased overhead which may be significant for short data frames.

Furthermore, the RTS/CTS mechanism can be regarded as a way to improve the MAC protocol performance. In fact, when the mechanism is enabled, collisions can obviously occur only during the transmission of the RTS frame. Since, the RTS frame is usually shorter than the data frame the wastage in bandwidth and time due to the collision are reduced.

In both cases the effectiveness of the RTS/CTS mechanism depends upon the length of the data frame to be "protected". It is reasonable to think that the RTS/CTS mechanism improves the performances when data frame sizes are large when compared to the size of the RTS frame. Consequently, the RTS/CTS mechanism relies on a threshold, the *RTS threshold*. The mechanism is enabled for data frame sizes over the threshold and disabled for data frame sizes under the threshold.

3. Simulation environment

The main difficulty for the analysis (at least) of the asynchronous part of the IEEE 802.11 MAC protocol via analytical models is the high degree of complexity and interdependence of the various processes that are involved in the protocol operation. Therefore, to evaluate the performances of the protocol we designed an *ad hoc* simulator and implemented it in C++ language.

3.1. Traffic characterization

Generally, when modeling data traffic, packet arrival processes are often assumed to be Poissonian. However, this model does not capture any correlation between consecutive packet arrivals, which is, on the other hand, exhibited by experimental data [10,12]. To recover from this problem MMPP (Markov Modulated Poisson Process) data traffic models have been proposed [7,10]. An MMPP process is characterized by an underlying Markov chain with N states. When the Markov chain is in state $\{i\}$ $(i = 1, 2, \ldots, N)$ the arrival process is Poisson with rate λ_i. A particular case of MMPP process is a two-state MMPP process where $\lambda_2 = 0$. Such a process is called the Interrupted Poisson Process and is basically an ON/OFF process.

One of the characteristics of MMPP processes is that the sojourn time in each state of the underlying Markov chain is exponentially distributed. However, recent teletraffic studies have shown that experimental data related to WWW applications can be satisfactorily modeled by ON/OFF processes where the ON and OFF time-length distributions are heavily tailed (e.g., Weibull, Pareto).[2]

Another issue in modeling data traffic is concerned with packet size distribution. In commonly used data applications (WWW, FTP, NNTP, Telnet, etc.), packets vary in size with a distribution which deviates considerably from the simple exponential distribution. Many studies have shown that the packet size distribution is bimodal (see, for example, [10]) as shown by the probability mass function (pmf) reported in figure 5, which was estimated by a real trace [1].

For the purpose of our simulation analysis we used the pmf shown in figure 5 to generate packet size. As far as

[2] As underlined in [15], by aggregating a large number of such ON/OFF sources, with infinite variance distributions for the ON and OFF time durations, the resulting traffic process exhibits self-similar properties.

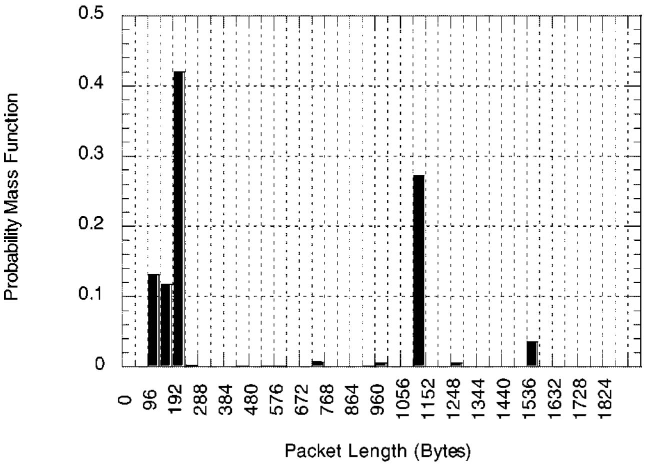

Figure 5. Packet size probability mass function.

the arrival process is concerned the following arrival models were considered:

1. Poisson.
2. ON/OFF where the ON and OFF periods are distributed
   - exponentially (MMPP),
   - according to Weibull.

In most of the analysis, for reasons explained in section 4, arrivals within ON periods are distributed exponentially.

It might be useful to recall that the Weibull distribution of a random variable X is

$$P\{X \leqslant x\} = F(x) = 1 - \mathrm{e}^{-(x/\beta)^{\alpha}} \quad \forall x \geqslant 0, \tag{1}$$

where $\alpha > 0$ and $\beta > 0$ are real numbers, and are called shape and scale parameters, respectively. Furthermore, the average $E[X]$ and variance σ_X^2 are

$$E(X) = \frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right), \tag{2}$$

$$\sigma_X^2 = \frac{\beta^2}{\alpha}\left\{2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha}\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2\right\}, \tag{3}$$

where

$$\Gamma(z) = \int_0^{\infty} t^{z-1}\mathrm{e}^{-t}\,\mathrm{d}t.$$

From (1) it is easy to check that when $\alpha = 1$ the Weibull distribution is reduced to an exponential distribution.

3.2. Performance measures

In this section we introduce the performance measures (or indices) used to characterize the quality of service (QoS) provided by the IEEE 802.11 MAC protocol. The indices defined in our analysis to specify the QoS are a widely accepted, minimum set of performance measures used to characterize the performance of any computer network.

Before introducing the performance measures we observe that packets are queued in a station *local queue* as soon as they arrive. Therefore, except when the station is empty, a packet will experience some delay in the local queue before contending for the channel according to the IEEE 802.11 MAC protocol. With this in mind, the performance measures we use in our analysis are:

- *average queuing delay*: average time elapsed from the time a packet joins the local queue *up until* it reaches the head of the local queue itself, i.e., it starts contending for the channel;
- *average MAC delay*: the average delay experienced by a packet from the time it reaches the head of the local queue up until the beginning of its successful transmission. The rationale behind this choice is that the average MAC delay measures the delay caused by the MAC protocol, and thus measures the interference between a user on one station and the users on the other stations. Furthermore, the average MAC delay does not include the average queuing delay experienced by a packet while queued in the local station queue, i.e., it does not take into account the interference between users on the same station;
- *average access delay*: sum of the average MAC delay and the average queuing delay in the local queue;
- *aggregate throughput*: average number of bits successfully transmitted by all stations per time unit.

Table 1
System parameter values.

System parameter	Parameter value (μsec)
Slot-time	50
SIFS	28
DIFS	128 (SIFS + 2 Slot-time)
Medium capacity	1 Mbit/sec

Table 2
Traffic parameter values.

Traffic parameter	Parameter value
Station Offered Load	30 Kbit/sec[1]
Average ON duration	3.3 sec
Average OFF duration	22.8 sec

[1] This is Offered Load of any station in the experiments discussed in section 4. This is not always true in the experiments reported in section 5 as specified in that section.

3.3. Operation parameter setting and assessment scenarios

The system parameters for our simulation environment are reported in table 1. These values are specified in the IEEE 802.11 standard [9]. Traffic related parameters are reported in table 2. Specifically, the station offered load value of 30 Kbit/sec was chosen to cover a broad range of data services while the ON and OFF duration, although different from the values in [5], allow reasonable execution times for simulation experiments without altering the stochastic nature of the arrival process.

4. Simulation results for the balanced load environment

In this section we report and comment on, in terms of MAC protocol mechanisms, the simulation results we obtained with the arrival processes and operation parameter settings previously specified. Furthermore, we report results which show the sensitivity of the IEEE 802.11 MAC protocol to larger cell coverage and higher channel speeds. In all the experiments discussed in this section all the stations are assumed to have the same Offered Load.

4.1. Influence of the arrival process with RTS/CTS disabled

In most of the curves reported below we compare Poisson, MMPP and ON/OFF processes where ON/OFF durations are Weibull distributed with several α values. Furthermore, we also varied the interarrival packet time distribution during ON periods. Specifically, we considered, in addition to the exponential distribution, the Weibull and constant distributions. The results obtained were substantially the same. For this reason, in most of the following experiments the interarrival packet time distribution is taken as being exponential.

Figure 6 shows the aggregate throughput achieved by the IEEE 802.11 versus the number of data sources. This figure highlights that the throughput curves which refer to the various arrival data processes are very close to each other.

The situation is very different from the average access delay standpoint shown in figure 7. This figure shows that the more the ON/OFF duration distributions deviate from the exponential one (i.e., the more heavily tailed the distribution is) the worse the average access delay. In order to understand the reasons for this behavior we measured both components of the average access delay. Figure 8 reports the average MAC delay and the average queuing delay versus the number of stations for three different values of α.

The above figure highlights that the average MAC delay remains constant in the range of α values we considered and this means that the collision avoidance mechanism is not influenced by the burstiness of the arrival process. By contrast, the average queuing delay increases significantly when the α values decrease. This behavior is certainly due to the fact that when the ON periods are distributed according to a heavy tail distribution the probability of the occurrence of long ON periods is not negligible and hence the probability of having a long local queue is not negligible either.

The influence of the α value on the system performances is now clear. In the following, in order to achieve reasonable execution times for simulation experiments, we will always use the exponential distribution (i.e., $\alpha = 1$) for the OFF and ON duration. In other words, the arrival process will be assumed to be MMPP.

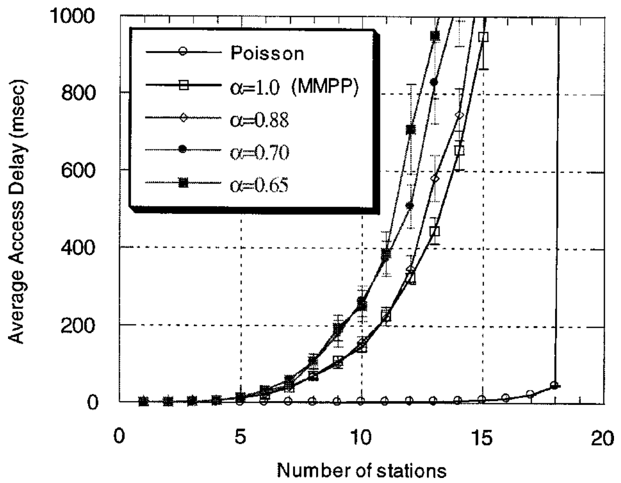

Figure 7. Average access delay versus the number of data sources for different arrival processes.

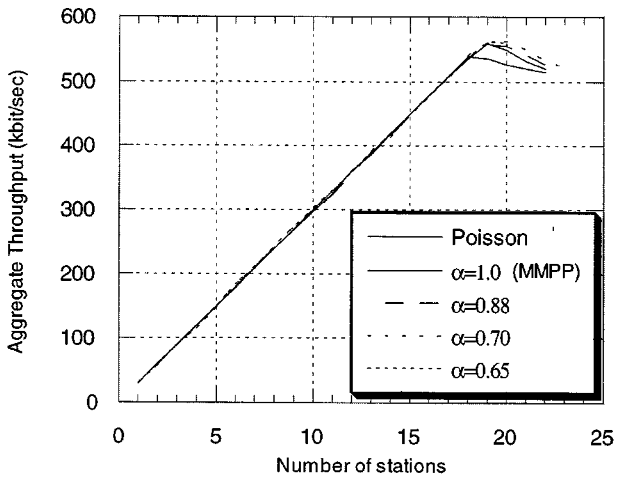

Figure 6. Aggregate throughput versus the number of data sources for different arrival processes.

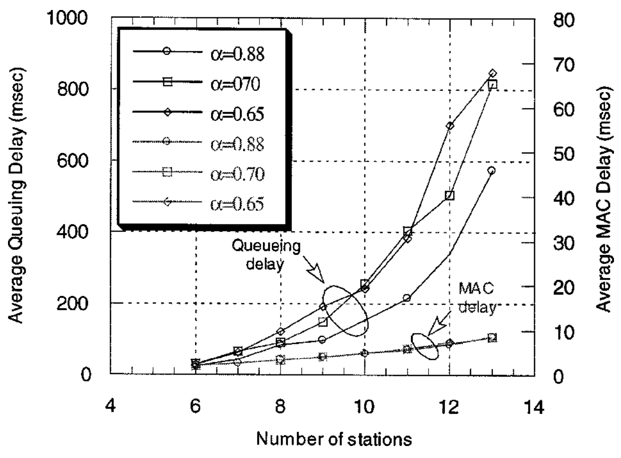

Figure 8. Average MAC delay and average queuing delay versus the number of stations for several values of α.

4.2. Influence of the arrival process with RTS/CTS enabled

In this section we look at the RTS/CTS mechanism which is here regarded as a means to improve the MAC protocol performances. The occurrence of collisions due to hidden stations is not taken into consideration. The purpose of the section is to analyze whether or not there exists an optimal RTS threshold for the type of traffic we consider. Figures 9 and 10 report the aggregate throughput and the average access delay vs the number of data sources in the cases in which the RTS/CTS mechanism is disabled and enabled and, in the latter case, for different RTS threshold values. Specifically, the values selected are 0 (which means RTS/CTS mechanism always enabled), 150 bytes and 280 bytes. This choice can be understood by making reference to the left bump in the pmf reported in figure 5. Obviously, 150 bytes falls in the middle and 280 on the right hand side of the above-mentioned bump.

As shown in figures 9 and 10, the RTS/CTS protocol mechanism significantly influences the performance measures for any RTS threshold value. This figure highlights two important aspects:

– the RTS/CTS mechanism improves the aggregate throughput and the average access delay with respect to the case in which RTS/CTS is disabled;
– when the RTS/CTS mechanism is enabled, the influence of the RTS threshold value both on the aggregate throughput and on the average access delay is almost negligible although it can be observed that the MAC protocol performances improve when the RTS threshold value decreases.

The latter point can be easily understood by taking into consideration that, due to the shape of the packet size pmf (see figure 5):

– by lowering the RTS threshold value, the number of packets affected by the RTS/CTS mechanism increases; however,
– the improvement in the RTS/CTS mechanism is very strong for packets belonging to the right bump (packets with very large packet sizes) while it is moderate or even low for packets falling in the left bump (packets with small to medium packet sizes). Hence, the improvement in the performance measures of the RTS/CTS mechanism is not significantly affected by the positioning of the threshold within the left bump.

To further deepen the influence of the RTS threshold value on the average access delay performance we investigated other packet length distributions. Specifically, for the uniform and exponential distributions with the same average as the distribution reported in figure 5, we observed the same behavior: i.e., the threshold value does not significantly affect the protocol performance.

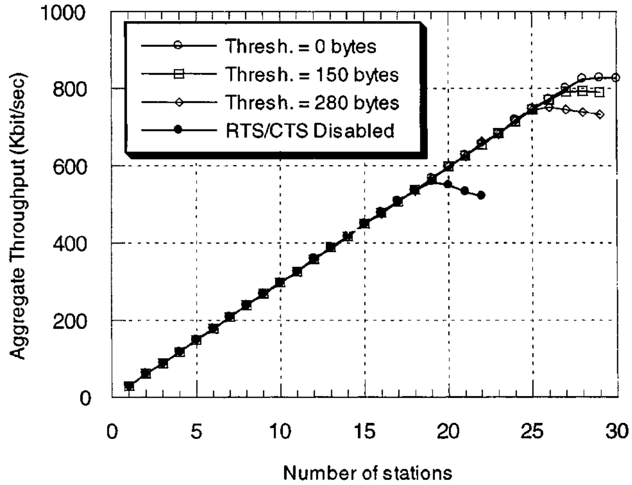

Figure 9. Aggregate throughputs when the RTS/CTS mechanism is enabled (for different threshold values) and disabled.

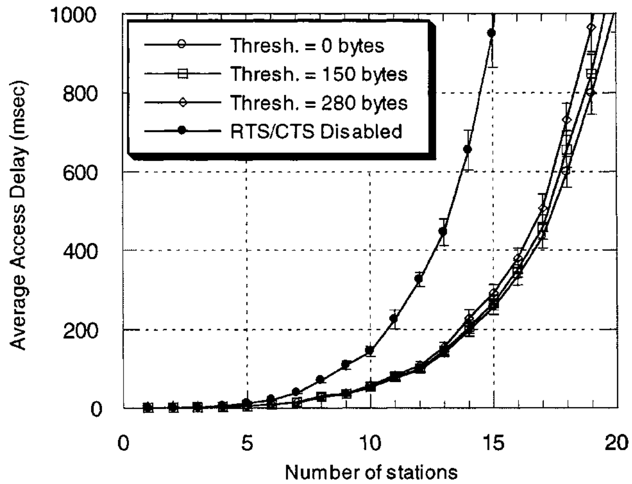

Figure 10. Average access delay when the RTS/CTS mechanism is enabled (for different threshold values) and disabled.

4.3. Influence of the slot-time duration

The analysis reported in this subsection is an extension of the analysis reported in [6]. As highlighted in section 2.2, the slot duration depends upon the maximum size of the cell coverage. Figures 11 and 12 report the throughput and average access delay versus the number of stations for several cell sizes expressed in terms of slot duration.

Figure 11 shows that the maximum achievable throughput decreases when the slot duration increases. Furthermore, from figure 12 it follows that the average access delay, for a given number of stations, increases when the slot duration increases. This can be explained by taking into consideration that an increase in the slot duration results in: (i) an increase in the DIFS and hence in the average backoff period, and (ii) an increase in the number of collisions, as shown in table 3.

The latter effect can be explained by considering that the vulnerable period (see figure 13) of the collision avoidance algorithm implemented by the IEEE 802.11 MAC protocol has a duration equal to the slot time. This can be understood

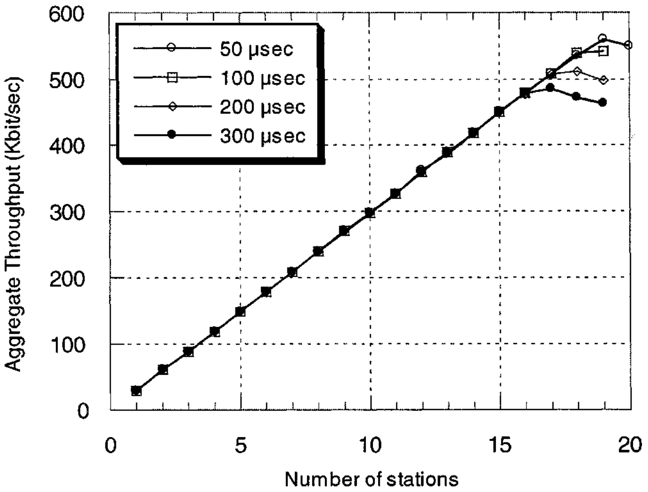

Figure 11. Influence of the slot-time duration on the aggregate throughput.

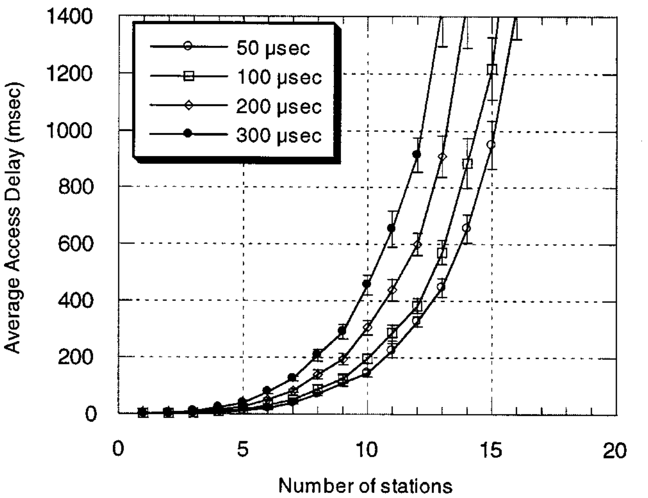

Figure 12. Influence of the slot-time duration on the average access delay.

Table 3
Mean number of collisions versus slot-time duration for three different numbers of stations.

Number of stations	Slot-time duration 50 μsec	100 μsec	200 μsec	300 μsec
13	0.19	0.20	0.23	0.26
16	0.28	0.31	0.35	0.50
19	0.56	0.81	1.05	1.12

by analyzing figure 13, which shows a scenario in which the backoff time of a station (called reference station) expires at time t_0. Due to the propagation delay, any other stations for which the backoff expires in between half of a slot before and after t_0 (stations 1 and 2, respectively – see figure 13) will collide with the reference station.

4.4. Analysis at higher speeds

As attention is now turning towards wireless LANs which can support several tens of Mbit/sec, we broadened

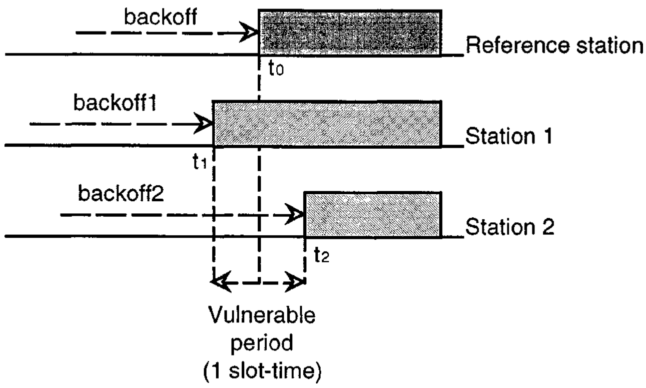

Figure 13. Vulnerable period.

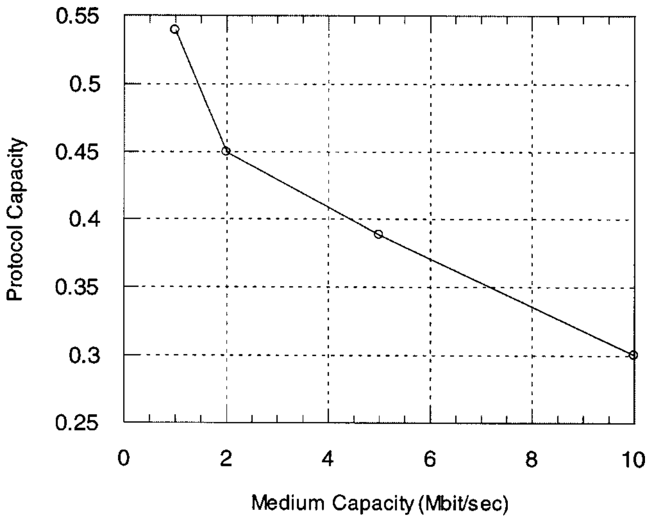

Figure 14. Protocol capacity versus different speed rates.

our analysis to verify whether or not the IEEE 802.11 MAC protocol is suitable for managing these rates efficiently.

We estimated the aggregate throughput versus the number of stations for channel speeds of 1, 2, 5, and 10 Mbit/sec. The results obtained show that the number of stations for which the maximum aggregate throughput occurs does not increase proportionally to the channel speed. This is due to the fact that when the channel speed increases, the packet transmission time decreases proportionally. However, since the slot time and, hence, the backoff time remain unchanged with the increase of the channel speed, the portion of time during which the channel remains unused due to the backoff algorithm is the same at any speed.

Figure 14 plots the IEEE 802.11 MAC protocol capacity (i.e., the maximum fraction of channel bandwidth used by successfully transmitted packets over all possible offered loads) for 1, 2, 5, and 10 Mbit/sec. As can be seen, the protocol capacity decreases when the channel speed increases. This implies that: (i) the IEEE 802.11 is not adequate to support the channel speeds planned for the future generation of wireless ATM LANs, and (ii) the protocol capacity depends upon the ratio (denoted by a in the literature – see [3]) between the packet transmission speed and the medium propagation time. The latter property should be expected since the IEEE 802.11 MAC protocol belongs to the class of random access MAC protocols which exhibit a similar dependency.

5. Simulation results for the unbalanced load environment

In all the previous experiments it was assumed that all the wireless stations have the same Offered Load, i.e., we considered a balanced distribution of the global load among stations. To make this scenario more realistic, we now assume that some stations are more loaded than others. To justify this choice, think for example about the Access Point of a BSS collecting all the traffic coming from any (wireless or wired) station outside the BSS itself and destined for the wireless stations in the BSS. In this example, the Access Point's Offered Load is usually many times greater than the Offered Load of a wireless station. This is especially true when the inter-BSS and/or internet components of the traffic are relevant with respect to the intra-BSS component of the total traffic. However, even assuming that all the traffic in a BSS is intra-BSS traffic (i.e., the wireless stations in the BSS transmit each other), in a local area environment, as the BSS is, some stations act as servers (e.g., WWW server, mail server, news server, file server, etc.) while others act as clients. Again, it is reasonable to think of these servers as having an Offered Load greater than the Offered Load of the other (client) stations accessing these servers.

To take the previous considerations into account we analyzed the performance of the IEEE 802.11 MAC protocol by considering an unbalanced distribution of the workload among the stations in the BSS. Specifically, we assumed that a particular station, hereafter *Server station*, has an Offered Load greater than the other stations referred to as *Client stations*. All the Client stations are characterized by the same value of the Offered Load. It is worth noting that the Server station considered here is not necessarily a server in the usual meaning of the term. For instance, it might be the Access Point of the BSS. By analogy, a Client station might be a wireless station transmitting to a destination outside the BSS.

The results obtained in the simulation experiments with unbalanced load distributions are reported and discussed in sections 5.1 and 5.2, respectively. The difference between the two sets of experiments is related to the number of stations and, hence, the amount of the aggregate Offered Load. In the set of experiments discussed in section 5.1 the number of active stations (Server + Clients) is constant and the aggregate Offered Load is kept constant as well. The analysis aims to investigate the influence on the performance measures of the percentage of traffic offered by the server station with respect to the (constant) aggregate Offered Load.

In the second set of experiments the Server Offered Load is a percentage of the traffic generated by all the Clients. All the active Client stations have the same Offered Load and their number is progressively increased in each experiment. The aim of this set of experiments is to investigate the influence on the performance measures of the number of active Client stations.

5.1. Constant global Offered Load

In the present set of experiments we started with a number $N_{st} = 15$ of wireless stations all having the same Offered Load (equal to 30 Kbit/sec). As shown in figure 6, with $N_{st} = 15$ the aggregate throughput is approximately half of the maximum achievable throughput. The average access delay experienced by each station in these conditions (dashed line in figure 15) can be derived from figure 7 by considering the proper arrival process (e.g., MMPP).

Of the above stations we then considered one particular station, the Server station, and, by keeping the total Offered Load constant, we progressively increased the fraction of traffic generated by the Server station while simultaneously reducing the fraction generated by all the other (Client) stations. The arrival process was assumed to be MMPP (i.e., $\alpha = 1.0$) both for the Server and for the Client stations.

The average access delays experienced by the Server and the Clients for different percentages of the Server Offered Load values are reported in figure 15. The dashed line reported in the same figure for the purpose of comparison represents the average access delay experienced by each station when the load is equally distributed among stations.

Figure 15 highlights that when one station is much more loaded than the others its average access delay becomes in the order of several seconds and dramatically increases as the imbalance in the distribution of the global Offered Load among stations increases. Even with a moderate imbalance (e.g., Server Offered Load of 20% in figure 15) the Server average access delay is approximately four times the average access delay experienced in the case of balanced load distribution among stations.

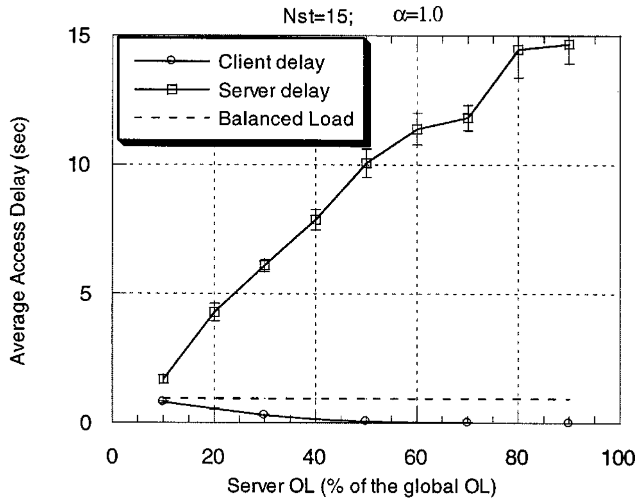

Figure 15. Client and Server average access delays for different percentages of the Server Offered Load.

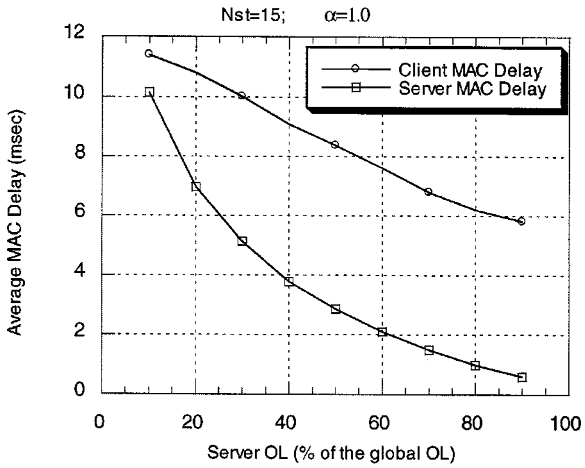

Figure 16. Client and Server average MAC delays for different percentages of the Server Offered Load.

On the other hand, the Client average access delay decreases as the percentages of the Server Offered Load increases.

Figure 16, which reports the average MAC delays components, suggests some additional and important considerations.

First, the average Server MAC delay is in the order of some milliseconds whereas the average Server access delay was in the order of some seconds. This means that the greatest contribution to the Server access delay is due to the queuing delay component and, as such, is related to the arrival process rather than the MAC protocol. By increasing the burstiness of the arrival process this component is expected to increase as experienced in section 4.1.

On the other hand, the average MAC delay is smaller for the Server than for the Clients. This can be explained by observing that most of the time the Server is the only station that has traffic to transmit. Hence, a percentage of Server transmissions are executed as immediate transmissions (i.e., without activating the backoff process). Even when the Server is involved in the backoff process (e.g., after a previous Server transmission) this process is not usually blocked by other transmitting stations. On the contrary, Clients almost always have to contend with the Server.

Obviously, as the imbalance in the global Offered Load increases, the probability that a Client has a frame to transmit decreases and, hence, the gap between the Server and Client average MAC delays grows, as shown in figure 16.

A further consequence of the above considerations is that the (Server and Client) average MAC delays decrease as the imbalance in the load distribution increases. In fact, if the probability that a Client has a frame to transmit decreases, the probability that a collision can occur decreases as well. Furthermore, the fewer the contending stations the less time actually needed by the backoff timer to expire (the backoff timer is frozen when another station is transmitting).

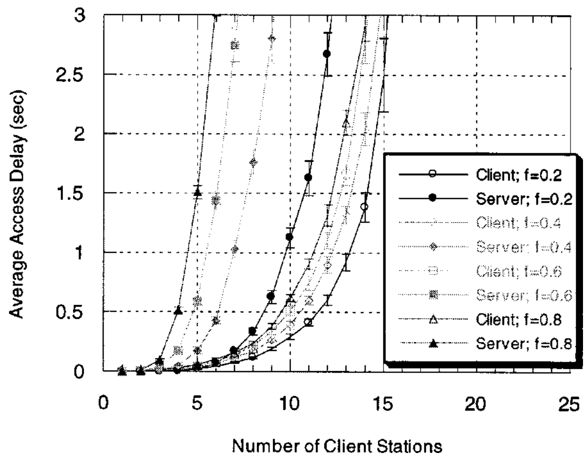

Figure 17. Average access delay versus the number of Client stations for several f values.

5.2. Variable Global Offered Load

The set of experiments discussed in the previous section was aimed at analyzing the performance of the IEEE 802.11 MAC protocol when one specific station is significantly more loaded than the others. However, the Server Offered Load was neither related to the number of the Client stations, which is fixed, nor to their Offered Load. If we look at a real wireless LAN environment what happens is that when a Client station becomes active, if the Client itself needs to communicates with the Server then the Server Offered Load increases. Obviously, the increase in the Server Offered Load depends on the particular type of communication between the Client and the Server, and, hence, in general, it varies from case to case. However, in the set of experiments we are going to discuss the following simplifying assumptions have been adopted:

(a) all the Client stations have the same Offered Load (equal to 30 Kbit/sec);

(b) the increase in the Server Offered Load due to a communication with a given Client station is a fraction f of the Client Offered Load;

(c) f is the same for each couple Client–Server.

Let OL_S and OL_C denote the Offered Load values for the Server and for a generic Client, respectively. Hence, if N_C ($N_C = N_{st} - 1$) indicates the numbers of active Client stations, from the above assumptions it immediately follows that $OL_S = N_C \cdot OL_C \cdot f$.

The average access delay experienced by each station versus the number of Client stations for several values of f is reported in figure 17.

Figure 17 highlights that there is a significant gap between the Server average access delay and the average access delay experienced by a Client station. Furthermore, the size of this gap increases as the number of Client stations increases and for a given number of Client stations it increases as the value of f increases. Even with a low number of stations (e.g., $N_{st} = 8$), despite the fact that the aggregate Offered Load is moderate (for $N_{st} = 8$ it is $OL < 400$ Kbit/sec even when $f = 0.8$) the Client average access delay is in the order of msec whereas the Server average access delay is in the order of some seconds.

6. Summary and conclusions

In this paper we have performed a simulation analysis of the IEEE 802.11 MAC protocol for wireless LANs. This analysis is based on "realistic" traffic models which are in line with those derived for modern data network applications (e.g., WWW). Both balanced and unbalanced load distributions among stations have been considered.

From our analysis with a balanced load distribution we can conclude that when the number of stations is in the order of 15, the IEEE 802.11 behavior is satisfactory. Under this condition, the IEEE 802.11 resource-sharing distributed algorithm results in an access delay experienced by each station in the order of one second (in our experiments) when the RTS/CTS is disabled, and approximately one third of a second when the RTS/CTS mechanism is enabled. This is a boundary which seems to be acceptable for today's data services.

Furthermore, at least for the type of traffic we consider, results obtained suggest using an RTS threshold equal to zero; i.e., to enable the RTS/CTS protocol mechanism for any packet length.

Finally, the IEEE 802.11 MAC protocol is not suitable for managing data traffic at channel speeds higher than a few megabit/sec.

In the unbalanced load distribution environment our analysis has pointed out that packets transmitted by a station much more loaded than the other ones suffer from a very large queueing delay. This may be in the order of several seconds even at moderate network loads.

Acknowledgement

The work described in this paper has been carried out under the financial support of the Italian Ministero dell'Università e della Ricerca Scientifica e Tecnologica (MURST) in the framework of the MOSAICO Project (Design Methodologies and Tools of High Performance Systems for Distributed Applications).

References

[1] Trace retrieved via anonymous ftp from `bellcore.flash.com`, file `/pub/lan_traffic/pAug.TL.` No longer available.

[2] H.S. Chhaya and S. Gupta, Performance of asynchronous data transfer methods of IEEE 802.11 MAC protocol, IEEE Personal Communications 3(5) (October 1996).

[3] M. Conti, E. Gregori and L. Lenzini, *Metropolitan Area Networks* (Springer, 1996).

[4] M.E. Crovella and A. Bestavros, Explaining World Wide Web traffic self-similarity, Technical Report TR95-015, Computer Science Department, Boston University (August 29, 1995).

[5] S. Deng, Empirical model of WWW document arrivals at access link, in: *Proceedings of ICC '96*.

[6] A. De Simone and S. Nanda, Wireless data: systems, standards, services, Wireless Networks 1(3) (October 1995) 241–254.

[7] H. Heffes and D.M. Lucantoni, A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, IEEE Journal on Selected Areas in Communications 4(6) (September 1986).

[8] ETSI, Hiperlan Functional Specification, ETSI Draft Standard (July 1995).

[9] P802.11/D2.1, Draft Standard IEEE 802.11, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (December 4, 1995).

[10] R. Gusella, Characterizing the variability of arrival processes with indexes of dispersion, IEEE Journal on Selected Areas in Communications 9(2) (February 1991).

[11] R.O. LaMaire, A. Krishna, J. Panian and P. Bhagwat, *Wireless LANs and Mobile Networking: Standards and Future Directions*, IBM Corporation (December 22, 1995).

[12] W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson, On the self-similar nature of ethernet traffic, IEEE/ACM Transactions on Networking 2(1) (February 1994) 1–15.

[13] M.A. Visser and M. El Zarki, Voice and data transmission over an 802.11 wireless network, in: *Proceedings of PIMRC '95*.

[14] J. Weinmiller, M. Schlager, A. Festag and A. Wolisz, Performance study of access control in wirelles LANs – IEEE 802.11 DFWMAC and ETSI RES 10 HIPERLAN, Mobile Networks and Applications 2(1) (1997) 55–68.

[15] W. Willinger, M.S. Taqqu, R. Sherman and D.V. Wilson, Self-similarity through high variability statistical analysis of Ethernet LAN traffic at the source level, in: *Proceedings of ACM SIGCOMM '95* (1995) pp. 100–113.

Giuseppe Anastasi received the Laurea cum laude degree in electronics engineering and the PhD degree in computer engineering, both from the University of Pisa, Italy, in 1990 and 1995, respectively. In 1991 he joined the Department of Information Engineering of the University of Pisa where he is currently an Assistant Professor of Computer Engineering. His main fields of interest include mobile computing and networking, distributed computing systems, modeling and performance evaluation of computer networks and distributed systems. He is a member of the IEEE Computer Society.
E-mail: anastasi@iet.unipi.it

Luciano Lenzini holds a degree in physics from the University of Pisa, Italy. He joined CNUCE, an institute of the Italian National Research Council (CNR), in 1970. Starting in 1973, he spent a year and a half at the IBM Scientific Centre in Cambridge, Massachusetts, working on computer networks. He has since directed several national and international projects including (in chronological order): RPCNET, the first Italian packet switching netwok; STELLA, the first European broadcasting satellite network and OSIRIDE, the first Italian OSI network. His current research interests include integrated service networks, the design and performance evaluation of both Metropolitan Area Network MAC protocols and packet-based radio access mechanisms for third generation mobile systems. He is author or co-author of numerous publications in these fields. He is currently on the Editorial Board of Computer Networks and ISDN Systems and a Member of the IEEE Communications Society. He has been on the program committees of numerous conferences and workshops, and served as chairman for the 1992 IEEE Workshop on Metropolitan Area Networks. In 1994 he joined the Department of Information Engineering of the University of Pisa as a Full Professor.
E-mail: lenzini@iet.unipi.it