Scheduling Time-sensitive Traffic on 802.11 Wireless LANs

Martin Heusse, Paul Starzetz, Franck Rousseau, Gilles Berger-Sabbatel, and Andrzej Duda

LSR-IMAG Laboratory
BP. 72, 38402 Saint Martin d'Hères, France
{heusse, starzetz, rousseau, gberger, duda}@imag.fr
http://www-lsr.imag.fr

Abstract In contrast to the common wisdom stating that 802.11 wireless LANs are not suitable for time-sensitive traffic, we have observed that in some conditions packet traffic transmitted over 802.11b may benefit from low delays even in saturation. Our analysis and measurements show that low delays can be obtained irrespectively of the greedy behavior of other hosts and without any traffic control mechanisms: when some hosts try to gain as much as possible of the transmission capacity of the radio channel, it is still possible for other hosts to experience low delay provided their packet rates are below some threshold value. The only situation in which a time-sensitive traffic source fails to obtain low delay is when its packet rate is too high with respect to its share of the channel capacity. We provide an analytical formula for determining the limiting packet rate that can be used to guide rate adaptive applications such as audio or video codecs to keep their output rates under the limiting rate and benefit in this way from low delays without any coordinated traffic control mechanisms.

1 Introduction

The common wisdom concerning time-sensitive traffic over wireless LANs such as 802.11 states that this kind of communication links cannot provide low latency [6, 11, 12]. Usually, it is assumed that the delay may be fairly long because of the Distributed Coordination Function (DCF) based on the CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance) medium access method. Such multiple access randomized protocols are considered as not suitable for time-sensitive traffic. Another access method defined in the 802.11 standard, the Point Coordination Function (PCF) oriented towards time-bounded services, is not implemented in most of current products. Many proposals try to alleviate this problem by modifying the MAC layer [2] [3] [7].

In this paper, we show that hosts generating time-sensitive traffic in a 802.11 cell may benefit from low delays even in saturation conditions. We observe that low delays can be obtained irrespectively of the greedy behavior of other hosts and without any traffic control mechanisms: even if some hosts try to gain as much as possible of the transmission capacity of the radio channel, other hosts may experience low delays provided their packet rates are below some threshold value.

We consider the case of several hosts that generate traffic over the wireless channel of 802.11. Some hosts send high priority time-sensitive flows that require low delay while generating packets with relatively small rate (for example H.261 typical rates start at 64 kbit/s and increase in multiples of 64 kbit/s). They compete for the radio channel with other hosts that do not care about the delay, but present a greedy behavior by trying to gain as much of the available bandwidth as possible.

In another work [10], we have proposed to use the DiffServ model to provide QoS differentiation at the IP level over the standard DCF method of 802.11. By scheduling packets according to their DiffServ classes (BE, AF, EF) and by constraining the output rate of each host via DiffServ traffic shaping mechanisms, we can keep the 802.11 network in the state of non-saturation so that the time critical high priority EF class benefits from stable short delays. Achieving such service differentiation requires traffic control mechanisms and collaboration of all hosts in a cell, for example a coordinator at the access point may configure the DiffServ mechanisms implemented in all hosts to reflect current allocations of the available bandwidth to aggregated traffic classes [8].

We begin with the analysis of the channel utilization in the 802.11 cell. The analytical results provide us with a limiting packet rate: if a host keeps its traffic below this limit, even if other hosts try to gain as much as possible, the host will experience short delays. We then verify experimentally this behavior. In our setup, we measure the throughput and the delay of two hosts in a wireless cell that generate traffic of different classes. We designate different traffic classes according the DiffServ model [5]: one host generates high priority EF traffic and the other one lower priority AF traffic. We use token buckets to control the source rates for which we want to measure the performance indices (they are only used for measurements, they are not needed for obtaining low delays). The experience confirms the analysis showing that if the channel is saturated by the lower priority AF class, it is still possible for the EF class to benefit from low delay provided that the EF packet rate remains lower than the limiting rate. The only situation in which the EF class fails to obtain low delay is when its packet rate is too high with respect to its share of the channel capacity. Note that the results of this paper can be used to guide rate adaptive applications such as audio or video codecs to keep their output rates under the limiting rates and benefit in this way from low delays without any coordinated traffic control mechanisms.

Our results show that the important parameter for scheduling traffic over the 802.11 WLAN is the packet rate and not the overall throughput. The importance of the packet rate results from the fairness properties of the CSMA/CA access method—in fact hosts in 802.11 share the channel capacity according to equal packet rates and not equal throughput shares [4, 9].

The paper is structured as follows. First, we analyze the utilization in 802.11b to derive the limiting packet rate (Section 2). Then, we describe the setup of the measurement experiments (Section 3). We show the performance results in a saturated cell (Section 4). Finally, we present some conclusions (Section 5).

2 Limiting packet rate in 802.11b

In this section, we model the behavior of a 802.11b cell [1] with hosts sending packets of different sizes to derive the limiting packet rate. The results of this paper follow up the analysis of the 802.11 performance anomaly [9], in which we have derived simple expressions for the useful throughput, validated them by means of simulation, compared with several performance measurements, and analyzed the performance of the 802.11b cell when one slow host (transmitting at a degraded rate e.g. 1 Mbit/s) competes with other fast hosts. Here, we modify the model to take into account different packet sizes and rates.

The DCF access method of 802.11b is based on the CSMA/CA principle in which a host wishing to transmit senses the channel, waits for a period of time (DIFS – Distributed Inter Frame Space) and then transmits if the medium is still free. If the packet is correctly received, the receiving host sends an ACK frame after another fixed period of time (SIFS – Short Inter Frame Space). If this ACK frame is not received by the sending host, a collision is assumed to have occurred. The sending host attempts to send the packet again when the channel is free for a DIFS period augmented of a random amount of time.

If there are multiple hosts attempting to transmit, the channel may be sensed busy and hosts enter a collision avoidance phase: a host waits for a random interval distributed uniformly over $\{0, 1, 2, ...CW - 1\} \times SLOT$. The congestion window CW varies between $CW_{\text{min}} = 32$ and $CW_{\text{max}} = 1024$, the value of $SLOT$ is 20 μs (these parameters are for 802.11b). The host that chooses the smallest interval starts transmitting and the others freeze their intervals until the transmission is over. When hosts choose the same value of the random interval, they will try to transmit at the same slot, which results in a collision detected by the missing ACK frame (only the transmitting hosts may detect a collision). Each time a host happens to collide, it executes the exponential backoff algorithm – it doubles CW up to CW_{max}.

We assume that each host i sends packets of size s_{i} at rate x_{i} packets per second. The frame transmission time depends on the size: $t_{\text{tr}} = s_{\text{i}}/R$, where R is the nominal transmission rate (11 Mbit/s for 802.11b). The overall frame transmission time experienced by a single host when competing with $N-1$ other hosts can be expressed as:

$$T_{\text{i}} = t_{\text{ov}} + \frac{s_{\text{i}}}{R} + t_{\text{cont}}.$$

where the constant overhead

$$t_{\text{ov}} = DIFS + t_{\text{pr}} + SIFS + t_{\text{pr}} + t_{\text{ack}}$$

is composed of the PLCP (Physical Layer Convergence Protocol) preamble and header transmission time $t_{\text{pr}} = 96\ \mu s$ (short PLCP header), $SIFS = 10\ \mu s$, t_{ack} is the MAC acknowledgment transmission time (10 μs if the rate is 11 Mbit/s as the ACK length is 112 bits), and $DIFS = 50\ \mu s$.

Under high load, to evaluate the impact of contention, we consider that the hosts always sense a busy channel when they attempt to transmit and that the number of transmissions that are subject to multiple successive collisions is negligible. In this case, we find:

$$t_{\text{cont}}(N) \simeq SLOT \times \frac{1 + P_{\text{c}}(N)}{N} \times \frac{CW_{\min}}{2},$$

where $P_{\text{c}}(N)$ is the proportion of experienced collisions for each packet successfully acknowledged at the MAC level ($0 \leqslant P_{\text{c}}(N) < 1$).

A simple expression for $P_{\text{c}}(N)$ can be derived by considering that a host attempting to transmit a frame will eventually experience a collision if the value of the chosen backoff interval corresponds to the residual backoff interval of at least one other host. Such an approximation holds if multiple successive collisions are negligible. So we have

$$P_{\text{c}}(N) = 1 - (1 - 1/CW_{\min})^{N-1}. \tag{1}$$

At this point we have all the elements of T_{i}, the global transmission time of host i. Now we want to find the overall performance—the channel utilization when hosts transmit packets at rate x_{i} while alternating transmissions. The utilization will determine the limiting packet rate beyond which the network enters the saturation state. We can evaluate the channel utilization by considering that host i uses the channel with rate x_{i} during time T_{i} as:

$$U = \sum_{i=1}^{N} x_{\text{i}} T_{\text{i}} + x_{\text{coll}}\ T_{\text{coll}}, \tag{2}$$

where $x_{\text{coll}}, T_{\text{coll}}$ are the collision rate and the time spent in collisions, respectively. If all hosts are greedy, their rates in the saturation state will be equal, so the limiting rate can be found from:

$$x^{\text{sat}} \sum_{i=1}^{N} T_{\text{i}} + x_{\text{coll}}\ T_{\text{coll}} = 1, \tag{3}$$

which finally yields:

$$x^{\text{sat}} = \frac{1 - x_{\text{coll}}\ T_{\text{coll}}}{\sum_{i=1}^{N} T_{\text{i}}}. \tag{4}$$

$x_{\text{coll}}, T_{\text{coll}}$ can be easily found for the case of two stations. To make the comparison with experimental results easier, we identify hosts by their type of traffic: host 1 generates time-sensitive EF traffic while host 2 generates AF packets:

$$x_{\text{coll}} = x^{\text{sat}} P_{\text{c}}(2), \tag{5}$$

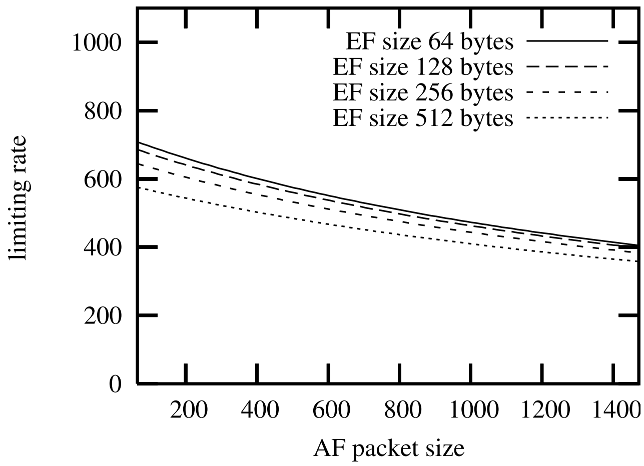

Fig. 1. Limiting packet rate for two hosts and different packet sizes.

$$T_{\text{coll}} = \max(T_{\text{EF}}, T_{\text{AF}}) \tag{6}$$

so for $N = 2$, and assuming that AF packets are longer or equal to EF packets, we obtain the following formula for the limiting rate:

$$x^{\text{sat}} = \frac{1}{T_{\text{EF}} + [1 + P_{\text{c}}(2)]T_{\text{AF}}}. \tag{7}$$

Figure 1 presents the limiting rate for two hosts in function of different packet sizes.

For $N > 2$, a simple approximation consists of not taking into account collisions. In this case we obtain the following upper bound for the limiting packet rate:

$$x^{\text{sat}} = \frac{1}{\sum_{i=1}^{N} T_{\text{i}}}. \tag{8}$$

3 Experimental setup

We have set up a platform to measure the delay and the throughput that hosts can obtain when sharing a 11 Mbit/s 802.11b wireless channel. We have used two notebooks running Linux RedHat 8.0 (kernel 2.4.20) with 802.11b cards based on the same chipset (Lucent Orinoco and Compaq WL 110). The wired part of the network is connected by an access point based on a PC box (SuSE 7.3) running software access point `hostap`. The notebooks use the Wvlan driver for the wireless cards. The cards do not use the RTS/CTS option that may optimize performance in case of the hidden terminal problem.

To avoid interferences in the use of the wireless channel, we measure the round trip time (RTT) in a configuration in which a host sends a packet over 802.11b

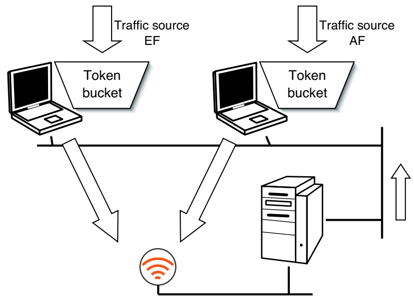

Fig. 2. Experimental setup.

and the reply returns via another interface (100 Mbit/s Ethernet). Figure 2 presents the experimental setup.

As said previously, we limit the experimental study to two hosts designated according to their type of traffic: the EF host sends time-sensitive traffic of a given packet rate whereas the AF host will try to increase its traffic as much as possible starting from 256 kbit/s to 10 Mbit/s in steps of 256 kbit/s.

The measurement results in the rest of the paper are presented in function of the *offered load*, which is the sum of the EF and AF traffic in kbit/s. The packet size given in figures corresponds to the UDP payload size.

4 Performance in saturation conditions

In this section we provide experimental results in saturation conditions. Figures 3, 4, 5 present the RTT of the EF class transmitting at different rates (128, 256, 512 kbit/s) when competing with the AF class. We can see that when the EF packet rate is small (128 kbit/s, 64 byte packets means 250 p/s packet rate), the RTT of the EF class remains small (under 6 ms) even if the cell is already saturated (offered load increased to 10 Mbit/s). As the limiting rate is 383 p/s for 1472 byte AF packets (cf. Eq. 7) and more for shorter packets, the EF class of 250 p/s packet rate will always benefit from low delays.

We can also see that for 512 byte AF packets and 256 kbit/s EF traffic (500 p/s packet rate), the delay is still short, because the limiting rate for this case is 644 p/s. Figure 5 shows the case in which the delay becomes very high due to queueing delays—for 1472 byte AF packets the limiting rate is 383 p/s, so the EF packet rate of 1000 p/s (512 kbit/s with 64 byte packets) is too high.

These result show that even if the channel is saturated by AF traffic, it is still possible for the EF class to benefit from low delay provided that the EF packet rate remains lower than the limiting packet rate. The reason for this behavior is the basic CSMA/CA channel access method which provides good fairness properties (contrary to the common wisdom concerning the fairness of 802.11 [4]) —the channel access probability is equal for all competing hosts.

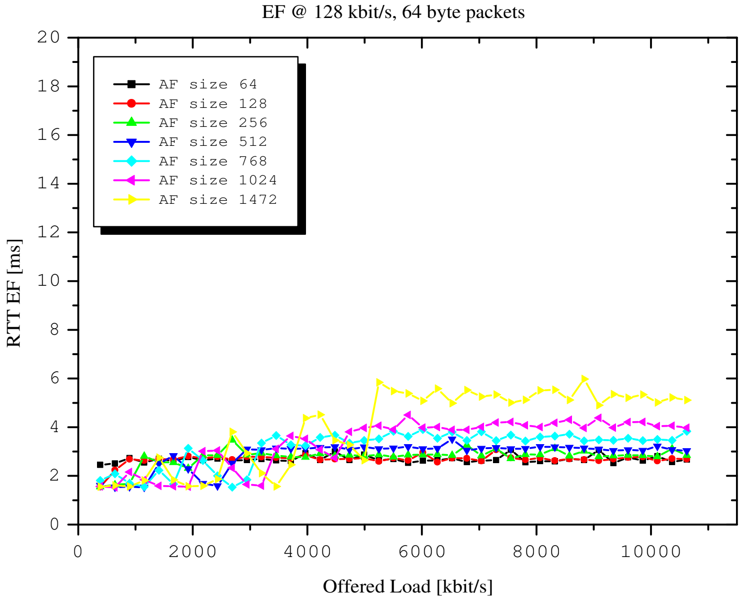

Fig. 3. RTT of the EF class for increasing offered load, constant 128 Kb/s EF traffic.

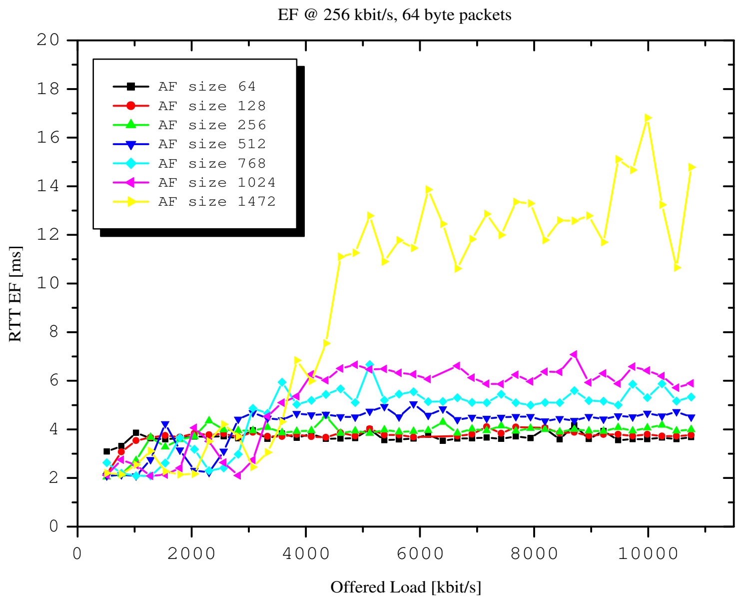

Fig. 4. RTT of the EF class for increasing offered load, constant 256 Kb/s EF traffic.

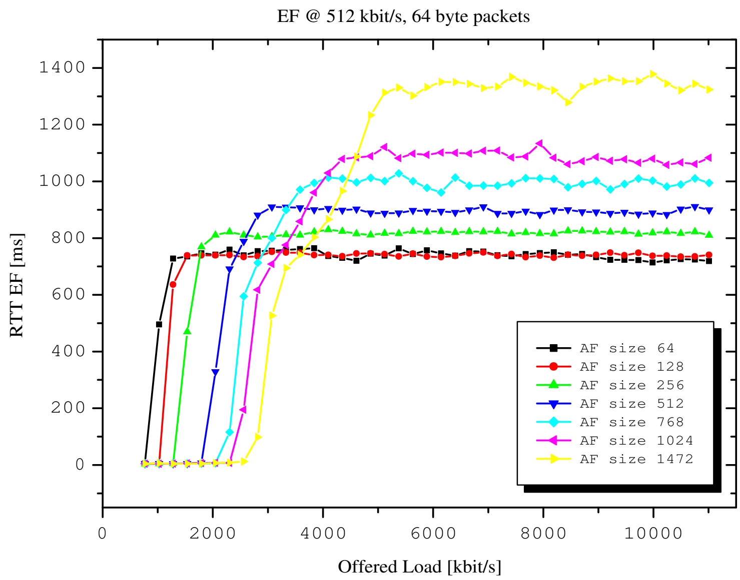

Fig. 5. RTT of the EF class for increasing offered load, constant 512 Kb/s EF traffic.

Hence at saturation, each competing class obtains an equal packet rate. And when the classes use different packet sizes, the throughput of each class may be different. So, even if the AF class sends packets with a rate exceeding its packet rate share, the EF class still benefits from its share of the packet rate. If the EF rate is lower than the packet rate share, its delay remains small.

The only situation in which the EF class fails to obtain a low delay is when the host packet rate is greater than its packet rate share. Figure 6 illustrates this case in a similar setup: for the constant EF rate of 1024 kbit/s, 64 byte packets, we increase the rate of the AF class for different packet sizes. This rate of the EF class corresponds to the packet rate of 2000 p/s, which is greater than the limiting rate for any AF packet size. We can observe from the figure that the EF class does not obtain this rate, so that the RTT increases because of queueing delays: the corresponding RTT measurements appear in figure 7. We can also observe how the packet rates of both classes tend towards equal values when the cell becomes saturated.

5 Conclusions

The analysis and measurements in this paper show that the time-sensitive EF class may benefit from low delays irrespectively of the greedy behavior of other hosts and without any traffic control mechanisms. The only condition for obtaining such desired behavior is to keep the packet rate under the limiting value

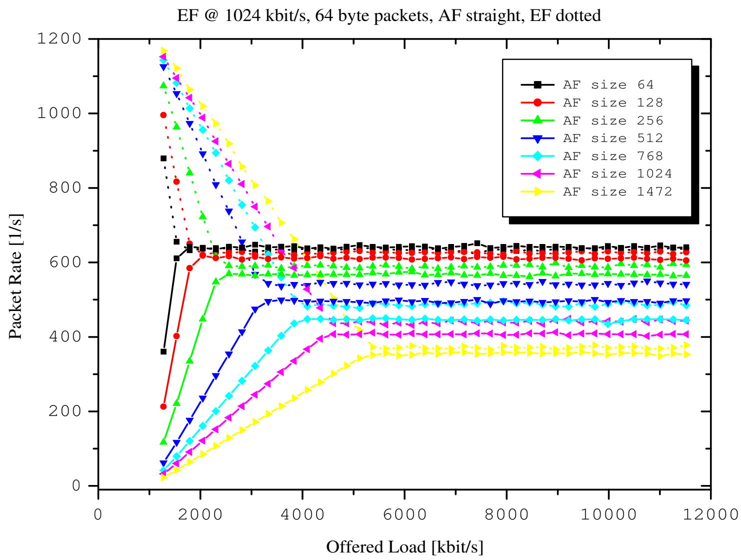

Fig. 6. Packet rates obtained by the AF and EF traffic for increasing offered load.

that we have analytically derived in Section 2. The only situation in which the EF class fails to obtain low delay is when its packet rate is too high with respect to its packet rate share.

Our results show that the packet rate is the most important parameter for QoS guarantees on the 802.11 WLAN. Its importance results from the fact that in the CSMA/CA access method, every time the EF host has a packet to transmit, it will contend with other hosts and gain the channel with probability $1/N$, where N is the number of hosts wanted to send a packet. If it does not succeed, it will attempt another time with a higher probability than the host that has gained the channel: its residual contention interval is smaller on the average than the contention interval of the successful host. Note also that our results apply to other variants of WLANs such as 802.11a and 802.11g, because they use the same MAC access method as 802.11b.

The results of this paper show that it is possible to provide some QoS guarantees over the standard DCF method of 802.11. In particular, adaptive applications such as audio or video codecs can keep their output rates under the limiting packet rate and benefit in this way from low delays without any coordinated traffic control mechanisms.

References

1. ANSI/IEEE. 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2000.

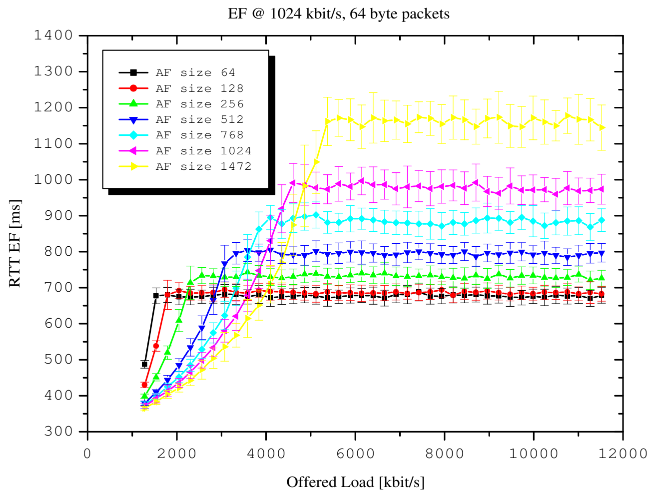

Fig. 7. RTT of the EF class, for an increasing offered load. High relative EF packet rate.

2. A. Banchs, M. Radimirsch, and X. Perez. Assured and expedited forwarding extensions for IEEE 802.11 wireless LAN. In *Tenth IEEE International Workshop on Quality of Service*, pages 237–246, 2002.
3. M. Barry, A.T. Campbell, and A. Veres. Distributed control algorithms for service differentiation in wireless packet networks. In *INFOCOM 2001*, volume 1, pages 582 –590, 2001.
4. Gilles Berger-Sabbatel, Andrzej Duda, Olivier Gaudoin, Martin Heusse, and Franck Rousseau. On the Fairness of 802.11. In *submitted for publication*, 2003.
5. S. Blake et al. An Architecture for Differentiated Services, 1998. `Internet RFC 2475`.
6. K.C. Chen. Medium access control of wireless LANs for mobile computing. *IEEE Network*, 8(5):50–63, 1994.
7. Garg, P. et al. Using IEEE 802.11e MAC for QoS over Wireless. In *IPCCC 2003*, Phoenix USA, 2003.
8. J. Antonio García-Macías, Franck Rousseau, Gilles Berger-Sabbatel, Toumi Leyla, and Andrzej Duda. Quality of Service and Mobility for the Wireless Internet. *Wireless Networks*, 9(4):341–352, 2003.
9. Martin Heusse, Franck Rousseau, Gilles Berger-Sabbatel, and Andrzej Duda. Performance Anomaly of 802.11b. In *Proceedings of IEEE INFOCOM 2003*, San Francisco, USA, 30– 3 2003.
10. Martin Heusse, Paul Starzetz, Franck Rousseau, Gilles Berger-Sabbatel, and Andrzej Duda. Bandwidth Allocation for DiffServ based Quality of Service over 802.11b. In *Proceedings of IEEE Globecom 2003*, San Francisco, USA, 2003.
11. L. Romdhani, Q. Ni, and T. Turletti. Adaptive EDCF: Enhanced Service Differentiation for IEEE 802.11 Wireless Ad Hoc Networks. In *WCNC'03*, New Orleans, USA, 2003.
12. J.L. Sobrinho and A.S. Krishnakumar. Real-time traffic over the IEEE 802.11 MAC. *Bell Labs Technical Journal*, 1(2):172–187, 1996.