

Understanding the Performance of 802.11 Networks

(Invited Paper)

Andrzej Duda

Grenoble Institute of Technology, LIG Laboratory

Email: duda@imag.fr

Abstract—In this paper, we review the most important performance characteristics of the 802.11 DCF wireless networks, point out some false common knowledge, and report on recent improvements.

I. 802.11 DCF BASICS

The IEEE 802.11 DCF (Distributed Coordination Function) [1] uses the *Carrier Sense Multiple Access/Collision Avoidance* (CSMA/CA) access method: a station trying to transmit first checks if the channel is free, waits for the DIFS interval (Distributed Inter Frame Space), and then transmits if the medium is still free. The receiving station sends an ACK frame after the SIFS interval (Short Inter Frame Space) if the frame is correctly received. When the station senses the channel busy, it waits until it is free and after the DIFS interval, it chooses random backoff b , an integer uniformly distributed in the contention window $[0, CW]$ ($CW_{\min} = 31$ for 802.11b and $CW_{\min} = 15$ for 802.11g) and counts down for b SLOT intervals before attempting to transmit. If another station transmits before the end of the backoff, the count down freezes and the remaining time is used in the next transmission attempt. If two stations have the same values of the backoff (or the remaining backoff), they transmit at the same instant and collide. A station can detect a collision, because it does not receive an ACK for its frame. In this case, it applies the *exponential backoff* algorithm—it doubles CW up to the maximal value of CW_{\max} . CW returns to the minimal value of CW_{\min} after a successful transmission.

This access method presents several advantages as well as drawbacks. It is fully distributed and allocates the channel to stations with roughly the same probability: if contending stations send frames of the same size and with the same bit rate, DCF allocates the same share of the channel capacity to all stations thus supporting *long-term fairness*. It achieves the best performance for a small number of stations with the optimal number depending on a given 802.11 variant, e.g. between 3 and 4 for 802.11b. However, if the number of contending stations increases beyond the optimal value, stations experience a significant collision rate, which lowers their performance. DCF also suffers from short-term unfairness for a greater number of stations: stations that collide increase their contention window and have less probability of accessing the channel. Exponential backoff implies that a station cannot distinguish between a collision and a lost frame due to bad channel conditions.

In this paper, we review the most important performance

characteristics of 802.11 DCF wireless networks by considering throughput, fairness, and delay for 802.11b and 802.11g variants.

II. THROUGHPUT

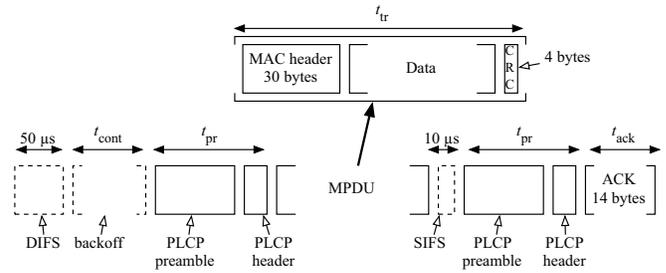


Fig. 1. Successful transmission of a single frame under 802.11b

The throughput of the 802.11 DCF depends on the bit rate and the access method overhead.

A. Access method overhead

Let us first consider the constant overhead of one data frame transmission (we do not take the random backoff into account yet). The overhead includes DIFS and SIFS intervals, ACK transmission as well as PLCP (Physical Layer Convergence Protocol) preamble and header (t_{pr}) that precede each frame. Figure 1 presents the time diagram of a single frame transmission and Table II gives the parameters of 802.11 variants.

TABLE I
802.11 MAC-PHY PARAMETERS

variant	bit rate (Mb/s)	DIFS	SIFS	SLOT	t_{pr} μ s
802.11b	1, 2, 5.5, 11	50 μ s	10 μ s	20 μ s	192, 96
802.11g	6, 9, 18, 24, 36, 48, 54	28 μ s	10 μ s	9 μ s	22.1

Let us denote by t_d , t_{tr} , t_{ack} the transmission time of the 1500 bytes data payload, the entire data frame (1534 bytes), and the ACK frame (14 bytes), respectively. The upper bound on efficiency U of 802.11b at 11 Mb/s bit rate is thus the following¹:

$$U = \frac{t_d}{\text{DIFS} + t_{pr} + t_{tr} + \text{SIFS} + t_{pr} + t_{ack}} = 0.79. \quad (1)$$

¹We consider the optional short preamble for this bound.

For 802.11g, the efficiency at 54 Mb/s becomes $U = 0.69$. Thus, a single station sending frames of 1500 bytes over 802.11b can at most obtain the throughput of 8.69 Mb/s and 37.26 Mb/s over 802.11g.

Fact 1: *The constant overhead lowers the throughput of the 802.11 DCF to about 20% or 30% of the bit rate depending on the variant.*

B. Random backoff and collisions

TABLE II
MAC-LEVEL THROUGHPUT, SIMULATION

N stations	1	2	4	10
802.11b, 11 Mb/s	7.16	3.78	1.88	0.70
cumulative	7.16	7.56	7.52	7.00
802.11g, 54 Mb/s	30.46	15.32	7.38	2.74
cumulative	30.46	30.64	29.52	27.74

When a station senses the channel busy, it waits for a random backoff, which further lowers its useful throughput (the random backoff even applies to a single station). On the average, a station waits $\bar{t}_{\text{cont}} = CW/2 \times \text{SLOT}$, which means that for a single station efficiency decreases to $U = 0.65$ for 802.11b (throughput of 7.15 Mb/s) and $U = 0.57$ for 802.11g (throughput of 30.68 Mb/s). Table II presents simulated throughput values for 802.11b and 802.11g in function of the number of contending stations (stations are greedy, i.e. they have always a frame of 1500 bytes to transmit). The cumulative throughput first increases (e.g. 7.56 Mb/s for $N = 2$ and 802.11b), because the interval between transmission attempts is shorter—it is the minimum of two or more uniformly distributed random variables. However, for a larger number of stations, the cumulative throughput decreases (e.g. 7.0 Mb/s for $N = 10$), because of an increased collision rate.

TABLE III
COLLISION RATE, SIMULATION

N stations	1	2	4	10
802.11b, 11 Mb/s	0.0	3.1	8.0	16.2
802.11g, 54 Mb/s	0.0	6.2	13.0	21.7

Table III shows simulated values of the collision rate that significantly increases with the number of stations. Note that it is much larger for 802.11g due to smaller CW .

Fact 2: *Random backoff increases the overhead (35% for 802.11b and 43% for 802.11g) and available throughput (7.15 Mb/s and 30.68 Mb/s). Increasing contention between stations results in collisions that further decrease the available throughput.*

C. Influence of transmission errors

The quality of received signals over a wireless channel changes in time due to multiple causes (noise, fading, interference, multipath propagation, station mobility) and may result in incorrectly received frames and decreased throughput.

Under 802.11 DCF, the influence of frame errors is even more important, because failed transmissions are undistinguishable from collisions: stations apply the exponential backoff after a lost frame. Table IV shows simulated values of throughput for a single station under 802.11b for different values of e , the frame error rate. It is lower than the throughput that could be obtained by a station without the exponential backoff after a frame loss: $U \times (1 - e)$.

TABLE IV
THROUGHPUT FOR 802.11B, FRAME ERRORS, $N = 1$, SIMULATION

e , frame error rate	1%	2%	5%	10%
throughput (Mb/s)	7.08	7.00	6.75	6.34
$U \times (1 - e)$ (Mb/s)	7.08	7.00	6.79	6.43

Fact 3: *Transmission errors result in a lower throughput amplified by the 802.11 contention control based the exponential backoff.*

D. Channel adaptation and rate diversity

A station may lower its bit rate at some threshold frame error rate to obtain better throughput—a station transmitting at a lower bit rate uses a more robust modulation scheme that decreases the frame error rate. Obviously, the transmission at the lower bit rate takes longer, but it is compensated by a lower frame error rate, which globally results in a better throughput. The main issue is to decide when it is beneficial to switch to a lower bit rate.

Let us consider a single station transmitting frames of the same size at the higher bit rate (lower bit rate) r_h (respectively r_l). When switching to the lower bit rate, we expect to lower the frame error rate from e_h to e_l . We assume that some part of the frame transmission time does not depend on the bit rate: σ_h (respectively σ_l) denote the proportion of useful throughput when using the higher bit rate (respectively lower bit rate). To find the threshold value of the frame error rate for which a station needs to lower its bit rate, we need to have at least the same throughput for both bit rates:

$$r_h \sigma_h (1 - e_h) = r_l \sigma_l (1 - e_l). \quad (2)$$

The threshold frame error rate is thus

$$e_h = 1 - \frac{r_l \sigma_l}{r_h \sigma_h} (1 - e_l). \quad (3)$$

As a first approximation, we can assume that lowering the bit rate will reduce the frame error rate to 0 and the proportions of useful throughput are equal ($\sigma_h = \sigma_l$). In this case, we simply obtain

$$e_h = 1 - \frac{r_l}{r_h} \quad (4)$$

This means that for 802.11b, a station needs to switch from 11 Mb/s to 5.5 Mb/s when the frame error rate exceeds 50%. If the frame error rate at the lower bit rate is not 0, this value should be even increased, e.g. if $e_l = 0.2$, the threshold increases to 60%.

Fact 4: Under imperfect channel conditions, switching to a lower bit rate is beneficial if the frame error rate exceeds some significant threshold.

E. Performance anomaly

Bit rate diversity in 802.11 leads to *performance anomaly*: the rate of a slower station limits the throughput of fast ones, because it takes more time to transmit frames at a lower bit rate, which decreases the channel time available for the transmission of fast stations [2].

To understand this effect, let us consider two stations transmitting at different bit rates: one station uses the high transmission rate $R = 11$ Mb/s and another one transmits at a degraded rate $r = 5.5, 2$, or 1 Mb/s. If both stations send frames of size s , they respectively spent $T_f = s/R + t_{ov}^R$ and $T_s = s/r + t_{ov}^r$ in transmission (t_{ov}^R and t_{ov}^r denote all overhead that does not depend on the bit rate: headers, interframe spaces, contention, and time spent in collisions). As the long-term channel access probability is equal for both stations, their mean throughput is as follows:

$$X_f = X_s = X = \frac{s}{T_f + T_s},$$

which means that both stations obtain the same throughput. We also have:

$$\frac{1}{X} = \frac{1}{R} + \frac{1}{r} + s(t_{ov}^R + t_{ov}^r), \quad (5)$$

which means that throughput roughly corresponds to a geometric mean of the respective bit rates. We provide more detailed analysis and results for any number of stations elsewhere [2]. The main consequence of this relation is that the slow station degrades the throughput of fast stations—they see their throughput reduced to the order of magnitude of the slow station throughput.

Fact 5: Stations switching to lower bit rates to adapt to bad channel conditions may significantly lower the throughput of stations that use higher bit rates.

III. FAIRNESS

Fairness can be defined at different time scales: *long-term fairness* guarantees that the probability of successful channel access observed on a long term converges to $1/N$ for N competing stations. An access method is *short-term fair* if channel allocation is fair over short time intervals. In this case, each station can expect to access the channel after a short time, which in turn results in short delays. Note that short-term fairness implies long-term fairness, but inverse does not hold.

A. Short-term fairness for small number of stations

Koksal et al. analyzed the short-term unfairness of an early Wavelan wireless card [3] and identified the short-term unfairness problem of its access method in which stations performed exponential backoff whenever the channel was sensed busy. Since this paper, many authors have stated that 802.11 suffers from short-term unfairness and referenced it as

the paper that proves the short-term unfairness of 802.11 [4], [5]. However, they have not realized that the access method of 802.11 has changed with respect to that of the Wavelan cards: in the 802.11 DCF [1], the exponential backoff is only applied after a collision. This misleading common wisdom has emerged from the confusion of two different access methods.

In fact, short-term fairness of the 802.11 DCF in the case of two stations is fairly good [6]. We evaluated it by considering patterns of transmissions and computing their average *Jain fairness index* in a window of an increasing size [3]. It is defined as follows: let γ_i be the fraction of transmissions performed by station i during window w ; the fairness index is the following:

$$F_J(w) = \frac{(\sum_{i=1}^N \gamma_i)^2}{N \sum_{i=1}^N \gamma_i^2}. \quad (6)$$

Perfect fairness is achieved for $F_J(w) = 1$ (as in TDMA) and perfect unfairness for $F_J(w) = 1/N$. We have normalized the window size with respect to the number of stations and computed the Jain index for windows multiple of N .

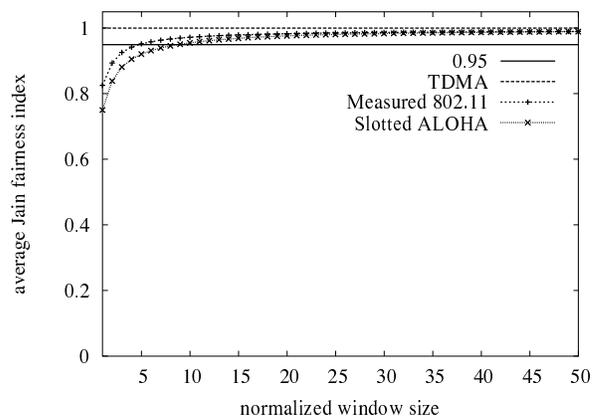


Fig. 2. Measured Jain fairness index for two 802.11 DCF stations

Figure 2 shows the Jain fairness index measured for two stations. It can be seen that the threshold value of 0.95 is quickly attained for the normalized window size of 5. The figure also presents the fairness of Slotted Aloha, a multiple access randomized protocol with good fairness properties that was used by Koksal et al. as a fairness gauge [3]—802.11 presents even better short-term fairness. Such a good behavior comes from the fact that 802.11 stations use their residual contention intervals—when a station chooses a long interval, it will wait during one or several turns, but then it will eventually succeed, because its backoff gets smaller at each turn. Note also that 802.11 presents much better short-term fairness than the Ethernet CSMA/CD known for its *capture effect* [7].

Fact 6: For a small number of stations, 802.11 benefits from good short-term fairness.

B. Short-term fairness for a larger number of stations

For a larger number of stations, the fairness index gets worse, because of the exponential backoff—after a collision,

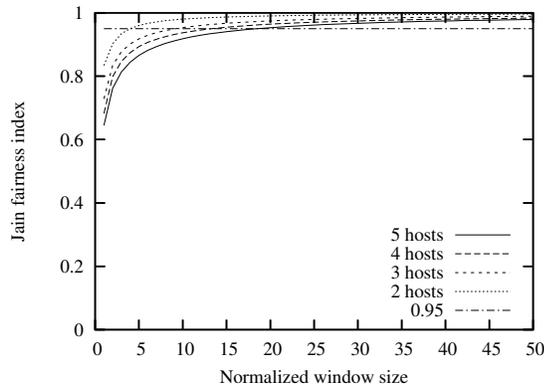


Fig. 3. Measured Jain fairness index of 802.11 for several stations

the stations that collided have more probability of choosing long backoffs, which gives other stations an increased transmission opportunity. Figure 3 shows the measured normalized Jain fairness index for several stations.

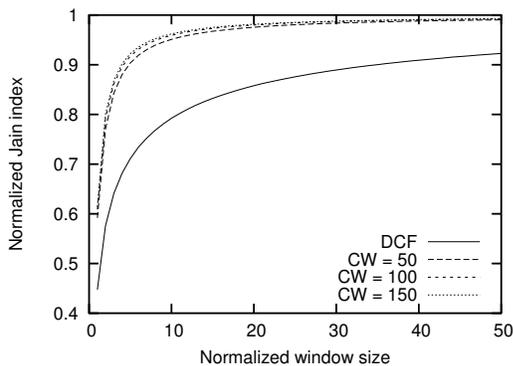


Fig. 4. Jain fairness index of 802.11 DCF with equal contention windows CW , five stations, simulation.

To evaluate the effect of the exponential backoff, Figure 4 compares the Jain fairness index for the 802.11 DCF with the case in which CW is kept equal for all stations. We can observe that equal CW results in significant improvement of short-term fairness.

Fact 7: *The short-term fairness of 802.11 becomes worse for an increasing number of stations due to the exponential backoff.*

C. Towards an optimal short-term fair access method

The previous example suggests that an access method needs to use the equal values of the contention window for all contending stations. For given traffic conditions (number of contending stations and their frame sizes), there is an optimal value of CW for which the stations benefit from short-term fairness and high throughput: if CW is too small, collisions are more frequent and if CW is too large, stations spend too much time waiting for transmission. Thus, an optimal access method should adapt the value of CW to the current traffic conditions.

To improve fairness and maximize throughput, we have proposed the *Idle Sense* access method [8]. In *Idle Sense*, contending stations make their contention windows dynamically converge in a fully distributed way to similar values solely by tracking the number of idle slots between consecutive transmissions. Each station measures n_i , the number of consecutive idle slots between two transmission attempts. Every $maxtrans$ transmissions, it estimates \hat{n}_i , the average of the observed values of n_i . Then, it uses \hat{n}_i to adjust its contention window to the target value n_i^{target} computed numerically for a given variant of IEEE 802.11 PHY and MAC parameters—its value is 5.68 for IEEE 802.11b and 3.91 for IEEE 802.11g [8]. When stations adjust their CW so that n_i converges to n_i^{target} , their throughput is optimal. Stations make n_i converge to n_i^{target} by applying AIMD (*Additive Increase Multiplicative Decrease*) [9] to contention window CW [10].

Figure 5 compares the measured Jain index of DCF and *Idle Sense* for five contending stations [10]. We can see that *Idle Sense* reaches the threshold value of 0.95 for a much smaller window size than the 802.11 DCF.

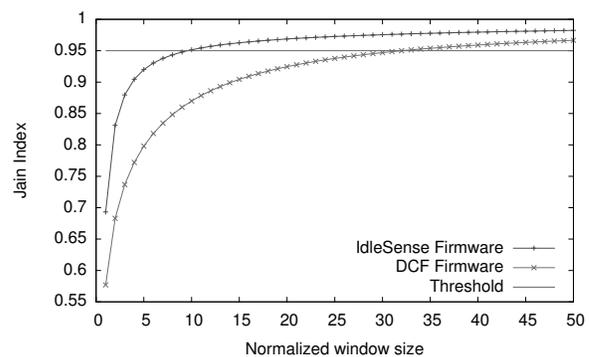


Fig. 5. Measured Jain index for 802.11 DCF and *Idle Sense* firmware, five contending stations.

Fact 8: *To improve short-term fairness, an optimal DCF-like access method needs to use the equal sizes of contention windows for all contending stations.*

D. Long-term fairness

When observed on a large time scale, the probability of channel access under 802.11 is equal for all stations, because most of the time they choose their random backoff from the same contention interval and when occasionally stations use larger backoffs after collisions, all stations have the same probability of colliding and deferring access attempts. If stations use the same data frame size, the equal channel access probability results in the equal throughput shares. Usually, such fairness is a desired property, however it is also the source of an important performance problem related to TCP.

In a 802.11 cell, an *access point* acts as a bridge for wireless stations: either it forwards frames between stations or interconnects the wired and the wireless parts of the network by forwarding data flows in two directions (upload or download) on behalf of wireless stations. As most of the

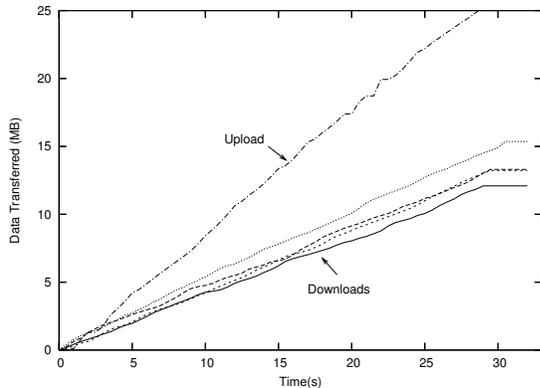


Fig. 6. Data transferred vs. elapsed time, 4 TCP downloads and 1 upload, IEEE 802.11 DCF. Experimental measurements.

traffic goes through the access point, it would require more channel capacity than wireless stations. Instead, if there are N wireless stations in a cell, the access point only benefits from $1/(N+1)$ of channel access probability, which leads to important performance degradation of TCP connections known as the *TCP unfairness problem* [11].

Consider wireless stations that communicate with stations on the wired part of the network through download or upload transport connections that go across the access point. Transport protocols usually use error control with one ACK per two data segments. In the pure upload case, stations have backlogged data segments to send, while the access point needs to send ACKs for each connection. Some ACKs may be lost at the access point due to insufficient share of the wireless link capacity and the limited size of buffers. However, TCP ACKs are cumulative, so that the upload throughput is only marginally affected by this situation.

For the pure download case, the situation is different with data segments backlogged at the access point. Because of the insufficient capacity share under IEEE 802.11 DCF, the segments of the download connection fill up the access point buffer and are dropped unless it is large enough for the TCP sources to choke on the window size [11]. This leads to important performance degradation: long delays, lost data segments, and retransmissions. Figure 6 illustrates this problem in a download oriented scenario: one upload competing with four downloads. We can observe that the uploading station obtains a far better throughput than the downloading ones.

We have proposed to solve the TCP unfairness problem in a simple and elegant way at the MAC layer [12]. We have defined the operation of an *Asymmetric Access Point* (AAP) that obtains transmission capacity that is k times greater than the capacity of *all* the wireless stations in the cell, i.e. the AAP benefits from kN times greater channel access probability compared to one wireless station, N being the number of active stations in the cell. Factor k corresponds to the number of data segments per ACK, which is 2 for most TCP implementations. The AAP achieves the allocation by applying the following surprisingly simple principles:

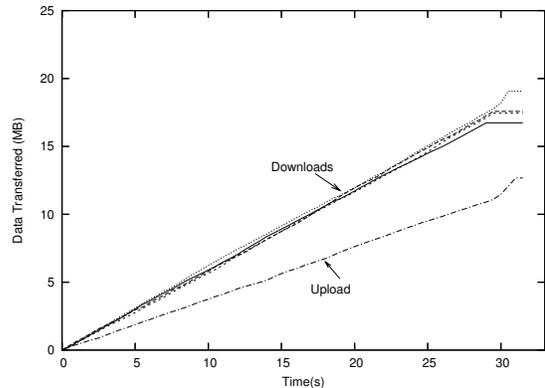


Fig. 7. Data transferred vs. elapsed time, 4 TCP downloads and 1 upload, AAP. Experimental measurements.

- 1) *The Asymmetric Access Point sets its contention window to a constant value independently of the number of active wireless stations.*
- 2) *Wireless stations use the Idle Sense access method [8] and its contention window adaptation mechanism.*

Figure 7 presents the corresponding throughput for the AAP scheme. We can observe that the uploading station benefits from less throughput and the performance of downloads is improved. By giving the access point a sufficient channel access share, AAP forces the downlink queue at the access point to be almost empty all the time so that TCP connections benefit from the desired operational state in which the destination controls the rate of the packet flow. Thus, instead of a low throughput due to congestion control and long delays caused by the saturated access point queue, we obtain significant performance improvement in terms of throughput and delay along with fair sharing of the wireless part of the network.

Fact 9: *Long-term fairness of 802.11 DCF results in significant unfairness at the level of TCP connections, because the access point does not benefit from sufficient capacity to convey download traffic.*

IV. DELAY

The common wisdom concerning time-sensitive traffic over the 802.11 DCF states that this kind of wireless LANs cannot provide low latency. However, stations generating time-sensitive traffic in a 802.11 cell may benefit from low delays even in saturation conditions [13], [14]. For a small number of stations, low delays can be obtained irrespectively of the greedy behavior of other stations and without any traffic control mechanisms: even if some stations try to gain as much as possible of transmission capacity, other stations can experience low delays provided their packet rates are below some threshold value.

Consider the case of two classes of stations: one sending time-sensitive traffic (EF DiffServ class) of a given packet rate whereas other stations (AF DiffServ class) increase their traffic as much as possible. An analysis of the channel utilization in this case yields a limiting packet rate: if the EF station keeps

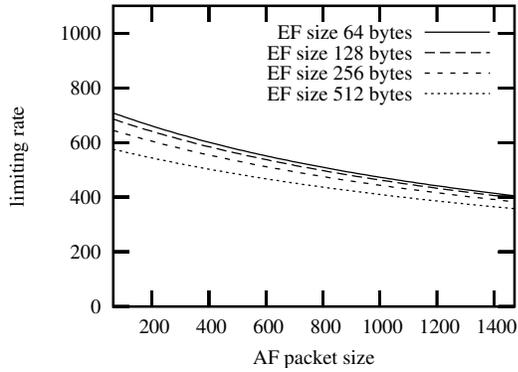


Fig. 8. Limiting packet rate for two stations and different packet sizes.

its traffic below this limit, even if AF stations try to gain as much throughput as possible, the EF station will experience short delays [14]. Figure 8 presents the limiting rate for two stations in function of different packet sizes.

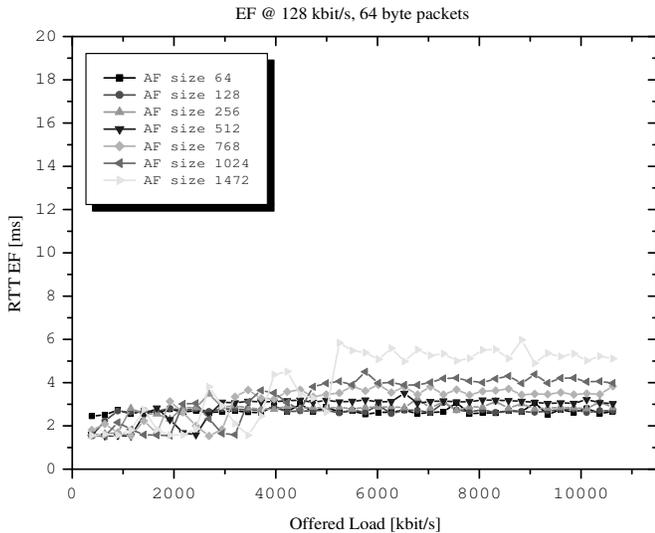


Fig. 9. Measured RTT of the EF class for increasing offered load, constant 128 Kb/s EF traffic.

Figure 9 presents the measured RTT of the EF class transmitting at 128 Kb/s (64 byte packets, 250 packet/s) when competing with the AF class. We can see that when the EF packet rate is under the limiting 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 and more for shorter packets, the EF class of 250 p/s packet rate always benefits from short delays.

Fact 10: For a small number of stations, a station contending with greedy ones can benefit from short delays provided its traffic stays under a limiting rate.

V. CONCLUSION

The 802.11 DCF is a simple random access method with an important protocol overhead that works fairly well for a small number of contending stations. In this case, it benefits from good short-term fairness and short delays mainly because of remaining backoff times used in channel contention. For an increasing number of stations, DCF suffers from much worse short-term unfairness, which results in longer delays. This effect comes from the exponential backoff that is not the best way of adapting contention to the number of stations. DCF allocates the channel in a fair manner, which is a suitable property for stations in the ad hoc mode, but it leads to the TCP unfairness problem in the infrastructure mode, because the access point suffers from insufficient share of channel capacity.

ACKNOWLEDGMENTS

Most of the results reported in this paper come from common work with my colleagues: Franck Rousseau, Martin Heusse, Paul Starzetz, Elena Lopez-Aguilera, Yan Grunenberger, Gilles Berger-Sabbatel, Romaric Guillier, and Olivier Gaudoin. This work was partially supported by the European Commission project WIP under contract 27402 and the French Ministry of Research project AIRNET under contract ANR-05-RNRT-012-01.

REFERENCES

- [1] IEEE, "IEEE 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1999.
- [2] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance Anomaly of 802.11b," in *Proc. of INFOCOM*, 2003.
- [3] C. Koksall, H. Kassab, and H. Balakrishnan, "An Analysis of Short-Term Fairness in Wireless Media Access Protocols," in *Proc. of ACM SIGMETRICS*, 2000.
- [4] C. L. Barrett, M. V. Marathe, D. C. Engelhart, and A. Sivasubramaniam, "Analyzing the Short-Term Fairness of IEEE 802.11 in Wireless Multi-Hop Radio Networks," in *Proc. 10th IEEE MASCOTS'02*, October 2002.
- [5] Y. Kwon, Y. Fang, and H. Latchman, "A Novel MAC Protocol with Fast Collision Resolution for Wireless LANs," in *Proc. of INFOCOM*, 2003.
- [6] G. Berger-Sabbatel, A. Duda, O. Gaudoin, M. Heusse, and F. Rousseau, "Fairness and its Impact on Delay in 802.11 Networks," in *Proc. of GLOBECOM 2004*, 2004.
- [7] K. Ramakrishnan and H. Yang, "The Ethernet Capture Effect: Analysis and Solution," in *IEEE 19th Conference on Local Computer Networks*, 1994.
- [8] M. Heusse, F. Rousseau, R. Guillier, and A. Duda, "Idle Sense: An Optimal Access Method for High Throughput and Fairness in Rate Diverse Wireless LANs," in *Proc. of ACM SIGCOMM 2005*, vol. 35, no. 4, August 2005, pp. 121–132.
- [9] D. Chiu and R. Jain, "Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks," *Journal of Computer Networks and ISDN*, vol. 17, no. 1, June 1989.
- [10] Y. Grunenberger, M. Heusse, F. Rousseau, and A. Duda, "Experience with an Implementation of the Idle Sense Wireless Access Method," in *Proc. of ACM CoNEXT 2007*, 2007.
- [11] S. Pilosof, R. Ramjee, D. Raz, Y. Shavitt, and P. Sinha, "Understanding TCP Fairness over Wireless LAN," in *Proc. of IEEE INFOCOM 2003*, vol. 2, March-April 2003, pp. 863–872.
- [12] E. Lopez-Aguilera, M. Heusse, Y. Grunenberger, F. Rousseau, A. Duda, and J. Casademont, "An Asymmetric Access Point for Solving the Unfairness Problem in WLANs," *IEEE Transactions on Mobile Computing*, 2008.
- [13] M. Heusse, P. Starzetz, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Bandwidth Allocation for DiffServ Based Quality of Service over 802.11b," in *Proc. of GLOBECOM 2003*, vol. 2, December 2003, pp. 992–997.
- [14] —, "Scheduling Time-Sensitive Traffic on 802.11 Wireless LANs," in *Proc. of QoFIS 2003*, LNCS 2856, Stockholm, Sweden, 2003.