An Analysis of the Backoff Mechanism used in IEEE 802.11 Networks

Marek Natkaniec and Andrzej R. Pach
Department of Telecommunications
University of Mining and Metallurgy
al. Mickiewicza 30, 30-059 Cracow, Poland
natkanie@kt.agh.edu.pl, pach@kt.agh.edu.pl

Abstract

This paper presents a simulation analysis of the backoff mechanism presented in the IEEE 802.11 standard. The obtained simulations allow us to determine the effective throughput and the mean packet delay versus offered load for different values of the Contention Window parameter and the number of contending stations. The choice of the CWmin and CWmax parameters was analyzed. In particular, the choice of CWmin value in dependence of the number of transmitting stations was presented. The simulation results show that the proper choice of the CW parameters has a substantial influence on the network performance.

1. Introduction

In last years, we can observe a permanent development of mobile devices such: laptops, palmtops and various pagers. These devices become independent computers with comparable functionality to desktop computers. It is necessary to support them with access to all types of networks. One has begun to equip these devices with suitable network cards assuring wireless access. The field of Wireless Local Area Networks (WLANs) is expanding very rapidly. WLANs permit for a large and free choice of architecture. They can be used as an extension of existing already wired networks. They can also be installed in places that are very difficult to wire as for example trading floors, manufacturing facilities, warehouses or historical buildings. The new possibilities of WLANs allow us to provide some new services. WLANs may be temporary or operational for short periods of time, when installation of wired networks is impractical. Two projects have been involved in standardizing the physical layer and the medium access control for WLANs, namely IEEE 802.11 [8] and ETSI HIPERLAN [7].

Media access protocols for a single shared wireless channel can be categorized as token-based or multiple access. Multiple access is a channel access protocol that is most frequently used. This kind of protocol compared to the token-based one is more robust. The nature of that protocol allows realizing bursty traffic (WWW, FTP, Telnet). The token-based protocols are better to support real-time, but less bandwidth required services (for example telephony or videotelephony). The second reason is in connection with handover procedures. Highly mobile terminals enter and leave the cells frequently. This would necessitate frequent token hand-offs or recovery in a token-based scheme. The more common medium access algorithm used is CSMA (Carrier Sense Multiple Access). Currently it is used in packet radio. In CSMA every contending station senses the carrier before the transmission. Carrier sense allows avoiding the collisions by testing the signal energy in the occupied band. The WLANs use a mutation of that algorithm called CSMA/CA (Carrier Sense Multiple Access Collision Avoidance). This algorithm has been employed by the DCF (Distributed Coordination Function) of IEEE 802.11. In order to avoid the collisions the random backoff mechanism has been used. It is based on a random slot selection basis from the Contention Window (CW) in which all stations participating in transmission are involved. The stations start to transmit their frames in random moments, to decrease the probability of collision. The detailed description of the random exponential backoff mechanism will be presented in the following.

The backoff mechanism has intensively been studied in the literature since the beginning of 70's. The idea of using the backoff mechanism in the MAC layer of the IEEE 802.11 standard has brought a new interest in such a mechanism. The proper selection of backoff parameters is an essential issue for the network performance. For example, the problem of unequal slot selection probabilities was considered in [13]. Two modified backoff schemes, namely weighted selection probabilities and load adaptive selection were proposed. These schemes can gain up to 20% in throughput and decrease the average access delay by 15%. The problem of CW parameter selection for Direct Sequence Spread Spectrum (DSSS) PHY was presented in [4].

The paper describes a simulation analysis of the backoff mechanism presented in the last version of the

IEEE 802.11 standard [8]. This analysis allows us to determine the realized throughput and the mean packet delay versus offered load for different CW parameters (CWmin and CWmax) and number of contending stations. Basing on the obtained results, it was possible to determine the optimal value of CWmin in dependence of the number of transmitting stations.

2. Distributed Coordination Function

The IEEE 802.11 standard supports two access methods: a mandatory Distributed Coordination Function (DCF) method which is available in both ad hoc and infrastructure configurations, and an optional Point-Coordinated Function (PCF) which is available in certain infrastructure environments and can provide time-bounded services [9], [10], [11].

DCF is the fundamental access method used to support asynchronous data transfer on the best effort basis. All the stations must support DCF. DCF employs the carrier sensing (CS) mechanism that check whether the signal energy in the occupied band does not exceed a given threshold so as to determine the medium is free and available for transmission. In order to minimise the probability of collisions a random backoff mechanism is used to randomise moments at which medium is tried to be accessed [1], [2], [3], [5], [12], [14].

The DCF protocol is enhanced further by provision of a virtual CS indication called Net Allocation Vector (NAV) which is based on duration information transferred in special RTS/CTS frames before the data exchange. It allows stations to avoid transmission in time intervals in which the medium is surely busy. The detailed description of DCF can be found in [8].

3. Exponential backoff mechanism

DCF adopts the slotted binary exponential backoff algorithm. A station desiring to initiate transfer of data shall utilize both the physical and virtual carrier sense functions to determine the state of the medium. If the medium is busy the station shall defer. After the DIFS time, the stations shall generate a random backoff period. This selected period should be deferred before transmitting. This procedure minimizes the probability of collisions during contention. The backoff time can be calculated from following formula:

$$\mathit{Backoff\_Time} = \mathrm{INT}\ (\mathit{CW} * \mathit{Random}\ (0,1)) * \mathit{Slot\_Time}$$

where, *Backoff_Time* is the time that must be deferred before the start of transmission; *Random (0,1)* is a pseudo random number between 0 and 1, drawn from the uniform distribution; *CW* is an integer between CWmin and CWmax; *Slot_Time* slot time duration; INT (X) is the integer part of X.

The CW parameter shall take an initial value of CWmin. It shall take the next value every time of unsuccessful attempt of transmission until reaching the value of CWmax. The CW shall remain at the value of CWmax until it will be reset. This improves the stability of the system in the case of heavy load. CW shall be reset to CWmin after every successful attempt of transmission. The set of CW values shall be sequentially ascending, integer powers of 2, minus 1, beginning with a CWmin and continuing up to CWmax value.

The standard specifies separately the values of CWmin and CWmax for each kind of physical layer. For Frequency Hopping Spread Spectrum (FHSS) the recommended values are following: CWmin=16, CWmax=1024 and SlotTime=50 µs. For DSSS we have the CWmin=32, CWmax=1024 and SlotTime=20 µs. For infrared physical layer (IR) the values are as follows: CWmin=64, CWmax=1024 and SlotTime=8 µs. The backoff procedure shall be invoked whenever a station desires to transfer a frame and finds the medium busy as indicated by either the physical or virtual carrier sense mechanism. A station performing the backoff procedure shall monitor the medium for physical activity during each backoff slot. If no physical medium activity is seen for the duration of a particular slot, then the backoff procedure shall decrement its Backoff_Time by a Slot_Time. If there is physical medium activity sensed at any time during a slot, then the backoff procedure is suspended, that is, the backoff timer shall not decrement for that slot. The medium shall be sensed as idle for the duration of a DIFS time, before the backoff procedure is allowed to resume. Transmission shall commence whenever the backoff timer reaches zero. The effect of this procedure is that when multiple stations are deferring and go into random backoff, then the station selecting the smallest Backoff_Time using the random function will win the contention. This situation is presented in Figure 1.

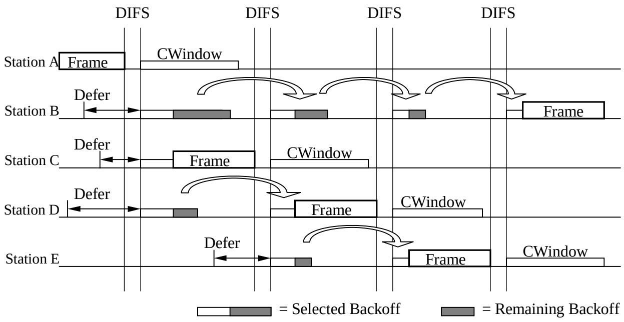

Fig. 1. Illustration of the Backoff Procedure

Considering the initial state it appears that each slot is selected with the same probability. In the next cycle all stations that competed for access reduce the backoff times. The new value is reduced by the time that elapsed until the wining station started its transmission. Within this reduced contention window all slots are selected with the same probability by the remaining stations. So, if a new station enters the competition or stations that collided in the previous cycle return back into it, they will choose slots within the whole range of contention window with the same probability [13].

Under high load there are always stations left in the competition as well as stations entering the competition. Then, we can notice then that the slots positioned earlier in CW have much higher probability to be chosen. It brings a very unprofitable effect. The slots that are chosen more likely are also chosen more likely again. This situation can bring a large number of collisions.

4. Simulation results

In order to investigate these phenomena an intensive simulations were performed. Several assumptions have been made to reduce the complexity of the simulation model:

- The effects of propagation delay are neglected. This is a very realistic assumption if the transmission distances between stations are of tens meters.
- The channel is error-free that means that each transmitted packet was successfully and correctly received at its destination.
- There are no stations operating in the “power-saving” mode. All stations are “awake” all the time. Then transmitted frames can be received immediately by the destination station.
- There is no interference from the nearby BSSs.

The gathered simulation results allow us to determine the realized throughput and the mean packet delay (as measured from the start of packet generation to the acknowledgement of its proper reception) versus offered load for different CW parameters. The frame was set to the constant size of 1050 bytes (overhead included). The next group of simulation results presents the realized throughput and the mean packet delay versus the number of stations for different values of the CW parameters. The last group of simulations allows us to determine the optimal value of CWmin in dependence of the number of contending stations.

DATA + ACK mode of transmission was used. The network was configured to 2 Mbps medium capacity. Almost all parameters were taken from the standard specification. The parameters used throughout all simulations are displayed in Table 1.

Table 1. Parameters used throughout all simulations

Parameter	Value
SIFS	10 μs
DIFS	50 μs
Length of ACK	14 octets
Length of slot	20 μs
Size of buffer for frames	1 frame
Number of retransmission of DATA frames	4
Timer 3	300 μs
Asynchronous data frame (headers included)	1050 octets
Medium capacity	2 Mbit/s
CWmin	variable: 8, 16, 32, 64, 128, 256, 512, 1024, 2048 or 4096
CWmax	variable: 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384 or 32768
Number of stations	variable: 10, 20, ... ,100

The results of obtained simulations are presented in three series of plots presented in Figures 2 - 10:

- The realized throughput (Fig. 2, Fig. 3, Fig. 4) and the mean packet delay (Fig.5, Fig. 6, Fig. 7) versus offered load for six different values of CWmin and CWmax and three different number of stations (5, 25 and 100).
- The realized throughput (Fig. 8) and the mean packet delay (Fig. 9) versus the number of stations for six different values of CWmin and CWmax and a large value of offered load, namely 4,5 Mbps (saturation of throughput).
- The realized throughput (Fig. 10) versus CWmin for 5, 25 and 100 stations (saturation of throughput).

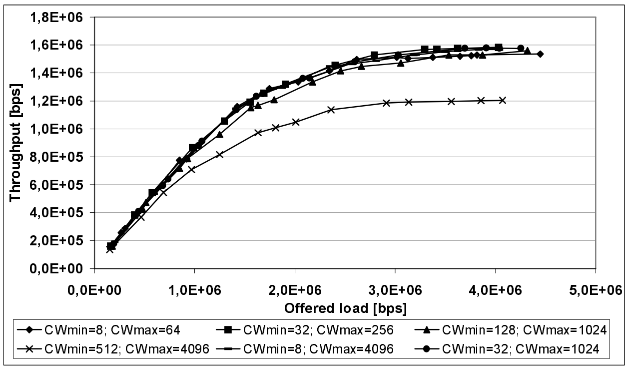

Fig. 2. Throughput versus offered load for 5 stations

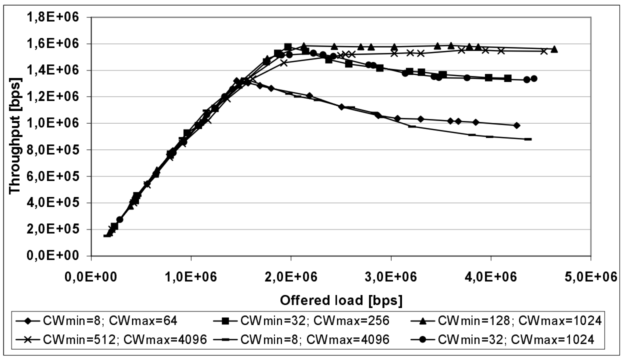

Fig. 3. Throughput versus offered load for 25 stations

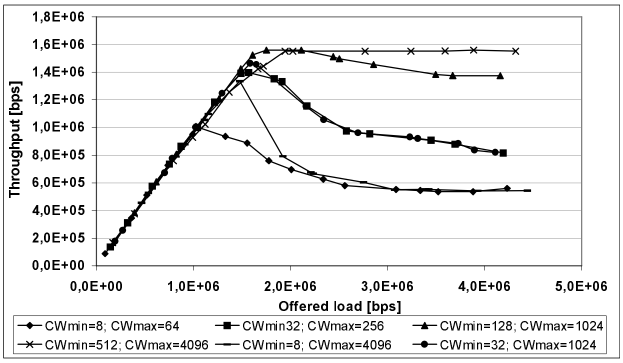

Fig. 4. Throughput versus offered load for 100 stations

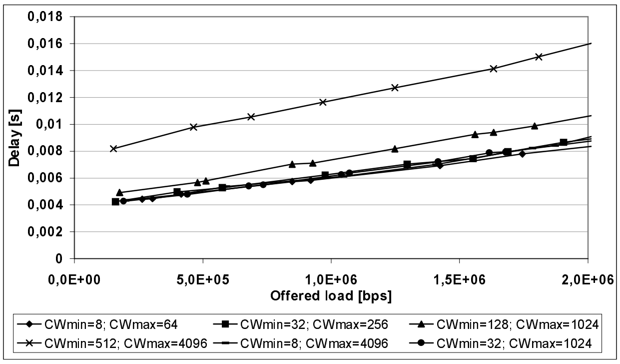

Fig. 5. Delay versus offered load for 5 stations

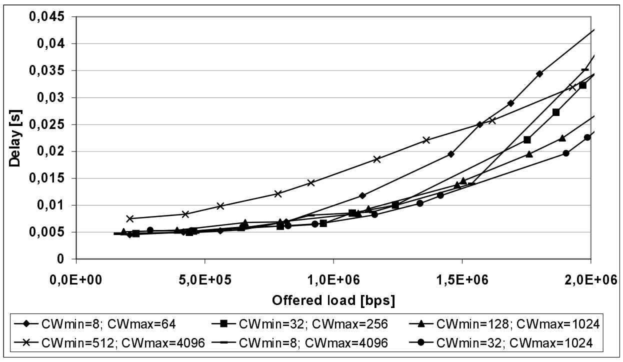

Fig. 6. Delay versus offered load for 25 stations

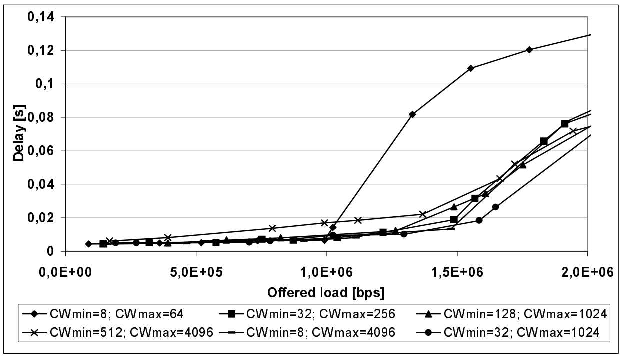

Fig. 7. Delay versus offered load for 100 stations

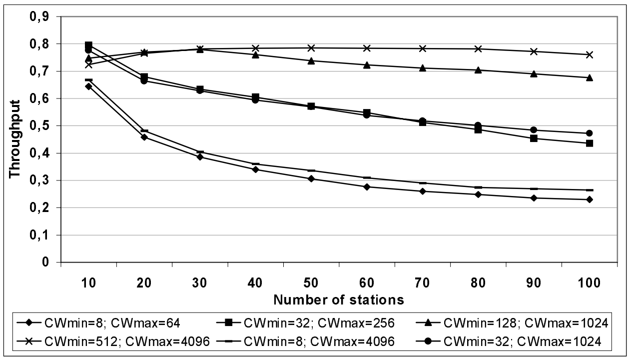

Fig. 8. Throughput versus number of participating stations

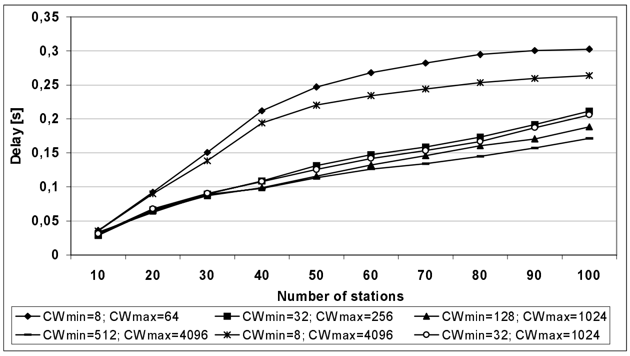

Fig. 9. Delay versus number of participating stations

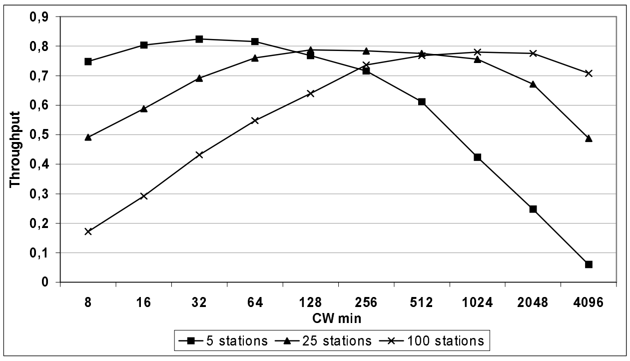

Fig. 10. Throughput versus initial CWmin size (CWmax=8*CWmin)

The detailed analysis of the results presented in figures allows us to draw several interesting conclusions. It can be observed that the proper choice of the CW parameters has a large influence on the network performance. This choice is highly dependent on the number of contending stations. The CW parameters varied in carried simulations as follows:

1) CWmin=8 and CWmax=64
2) CWmin=32 and CWmax=256
3) CWmin=128 and CWmax=1024
4) CWmin=512 and CWmax=4096
5) CWmin=8 and CWmax=4096
6) CWmin=32 and CWmax=1024

The simulation results obtained for 5 stations are presented in Figure 2. The worst performance is obtained for the maximum values of CWmin and CWmax, namely for CWmin=512 and CWmax=4096. This situation can be explained by the fact of too large values of Backoff_Timer even if we chose the slots from CW=CWmin. For 5 stations the proper selection of CWmin parameter (up to 128 slots) has less importance. The results obtained for 25 contending stations are presented in Figure 3. In this situation the best results are achieved for CWmin=128 and CWmax=1024. Slight worse results are achieved for CWmin=512 and CWmax=4096, especially for low values of network load. The decrease of throughput of about 25% for CWmin=32, CWmax=256 and CWmin=32, CWmax=1024 is observed. The worst situation is, as in case above, for CWmin=8, CWmax=64 and CWmin=8, CWmax=4096. Too large value of CWmax parameter brings performance degradation when a few stations are contending at large values of offered load. It is caused by too large size of CW. The stations that decrement their Backoff_Timers and no one send data cause the waste of a lot of bandwidth. There are some maxima of throughput especially observed for lower values of CWmin. Too small values of CWmin cause degradation of the network performance when the offered load increases (the number of collision increases). For 100 contending stations the best results are achieved for CWmin=512 and CWmax=4096 (see Figure 4). A little degradation of performance (about 20% of realized throughput) bring the choice of CWmin=128 and CWmax=1024 parameters. We got large degradation of throughput when using all other parameters. The worst situation is when we select CWmin=8 and CWmax=64. In this case the maximum number of slots in CW is less than the number of contending stations. This results in a large number of collisions and degradation of the network performance. So, we can draw the conclusion that proper selection of CWmin parameter has the greatest influence on the network performance. We should make an assumption that we do not reduce the CWmax parameter intentionally to a very low value. Too small value of CWmax can also bring the network degradation as shown above.

The mean delays for 5 stations are presented in Figure 5. The graphs can be approximated by linear functions. The biggest delay (2 times more than for the smallest values) is observed for CWmin= 512 and CWmax=4096. This is caused by a wrong selection of CW parameters (too large value of CWmin=512 slots compared to 5 contending stations). For CWmin=128 the mean packet delays are substantial bigger than for other values of CWmin. The mean delays as a function of offered load for 25 stations is presented in the next figure. The explicit fall of mean delays (about 3 times) for all values of offered load is observed. This steams from the fact of the smaller number of contending stations. A little bit greater delay at low values of offered load is observed for CWmin=512 and CWmax=4096. The mean packet delays versus offered load (up to 2 Mbps) for 100 contending stations are presented in Figure 7. The rapid growth of delay for above 1 Mbps of offered load for CWmin=8 and CWmax=64 is characteristic. For 2 Mbps it amounts 0,13 s (twice more than for other delays). It results from a large number of collisions and retransmissions of packets. All others plots are similar and the mean packet delay reaches the same values. A little bit greater delay, at small values of offered load, for CWmin=512 and CWmax=4096 is observed. This results from too large value of timeout at low values of offered load and large size of CWmin.

The Figure 8 allows us to answer the question about the dependence between the number of stations and the realized throughput for different values of CWmin and Cwmax. It is better to select less values of CWmin for smaller number of stations (CWmin=32 is the optimal value for 10 stations). The selection of too small value of CWmin causes degradation of the performance. For a larger number of stations it is better to select greater values of CWmin. For 20 stations, for example, the choice of CWmin=128 or CWmin=512 compared to CWmin=32 brings the increase of realized throughput by 10%. It is better to use CWmin=512 (maximum size of CWmin in our simulations) when there are more than 30 stations.

The mean delay versus number of stations is presented in Figure 9. In case of 10 stations the delay is of the same magnitude for all selected values of CW. An increased number of stations brings the growth of delay and it is the largest for CWmin=8, CWmax=64 (0,3s for 100 stations) and for CWmin=8, CWmax=4096 (0,27s for 100 stations). The mean delay is significantly less for all others CW parameters and in the case of 100 stations it reaches the smallest value equal to 0,175s for CWmin=512 slots. It reaches 0,18s for CWmin=128 slots.

Figure 10 presents a summary of the carried investigations. The dependencies between throughput, CWmin size and number of stations are presented. We assumed CWmax=8*CWmin (similar as specified in the standard). The observations show us that for 5 stations the optimal value of CWmin is 32 slots, for 25 stations – CWmin=128 slots and for 100 stations – CWmin=1024. The realized throughput is of 0,8 in all three cases.

5. Conclusions

This paper presents a simulation analysis of the backoff mechanism described in the IEEE 802.11 standard. DCF is the fundamental access method used to support asynchronous data transfer on best effort basis. To avoid the collisions the random backoff mechanism is employed. The random slot selection from the Contention Window (CW) by all transmitting stations is used. The stations start to transmit their frames in random moments, so the probability of collision decreases. The IEEE 802.11 standard allows us to use one of three different physical layers: FHSS, DSSS and IR. Standard specifies separately the values CWmin and CWmax for every kind of physical layer. At the beginning the standard recommended to use following dependence: CWmax=8xCWmin. However, this value is much larger in the latest version of standard. For FHSS the proposed values are: CWmin=16 and CWmax=1024, SlotTime=50 μs. For DSSS the proposed values are: CWmin=32 and CWmax=1024, SlotTime=20 μs. For infrared physical layer the values are as follows: CWmin=64 and CWmax=1024, SlotTime=8 μs. The standard does not specifies the dependencies between number of stations and CW parameters. The presented paper shows us those dependencies and allows us for network optimization to improve the performance. The simulation results show that the proper selection of parameters of backoff mechanism has a very large influence on the network performance. The wrong selection of CW parameters cause degradation of the throughput of 60% and the mean packet delay can grow several times. The proper selection of CW (CWmin and CWmax parameters) in dependence on number of participating stations seems to be the crucial point.

The presented analysis led us to some important conclusions, briefly presented below:

- Using too small values of CWmin at large number of stations brings a large number of collisions and degradation of network performance.
- Using too large CWmin at small number of stations brings degradation of the network performance. It is caused by too large size of CW. The stations that decrement their Backoff_Timer and no one sending the data wastes a lot of bandwidth.
- Using too small values of CWmax at large number of stations for large values of offered load brings a large number of collisions and degradation of the network performance.
- The most important issue is the correct choice of CWmin parameter. The choice of CWmin in dependence on number of contending stations is presented in Figure 10.
- The influence of CWmax parameter on the network performance is little. The CWmax parameter can not be reduced to a very low value, especially when there is a large number of contending stations.

References

[1] Bharghavan V., Demers A., Shenker S., Zhang L., *MACAW: A Media Access Protocol for Wireless LAN's*, ACM Sigcomm 1994.

[2] Buchholz D., *Comments on CSMA*, IEEE 802.11 Working Group paper IEEE 802.11-91/49.

[3] Chaya H. S., Gupta S. *Throughput and Fairness Properties of Asynchronous Data Transfer Methods in the IEEE 802.11 MAC Protocol*, Proc. of IEEE PIMRC '95, Toronto, September 1995.

[4] Diepstraten W., *Proposed values for Contention Windows (CW)*, IEEE 802.11 Working Group paper IEEE 802.11-95/80, May 1995.

[5] Diepstraten W., Ennis G., Belanger P., *DFWMAC - Distributed Foundation Wireless Medium Access Control*, IEEE 802.11 Working Group paper IEEE 802.11-93/190, Nov. 1993.

[6] Geiger E., *Making Sense Out of MAC and PHY Timing Specifications*, IEEE 802.11 Working Group paper IEEE 802.11-95/039, March 1995.

[7] HIPERLAN - High Performance Radio Local Area Network*; Functional Specification.* Draft ver. 1.1 ETSI January 1995.

[8] IEEE 802.11 Standard for Wireless LAN*; Medium Access Control (MAC) and Physical Layer (PHY) Specification*, New York Approved 26 June 1997.

[9] Natkaniec M., Pach A. R: *Simulation Analysis of Multimedia Streams Transmission in IEEE 802.11 Networks.* – ISWC'99 IEEE International Symposium on Wireless Communications, June 1999, Victoria, Canada.

[10] Natkaniec M., Pach A. R.: *Simulation Analysis of Service Integration in IEEE 802.11 Networks.* 5th Polish Teletraffic Symposium, April 1998, Warsaw, Poland.

[11] Natkaniec M., Pach A. R.: *A Performance Study of Point Coordination Function Efficiency of the IEEE 802.11 DFWMAC Protocol*– 6th Polish Teletraffic Symposium, April 1999, Szklarska Poręba, Poland.

[12] Rypinski C., *Limitations of CSMA in 802.11 Radiolan Applications*, IEEE 802.11 Working Group paper IEEE 802.11-91/46a.

[13] Woesner H., Weinmiller J., Wolisz A., *Modified Backoff Algorithms for DCF*, IEEE 802.11 Working Group paper IEEE 802.11-95/183, July 1995.

[14] Zegelin Ch., Kawaguchi D., *SIFS, PIFS, DIFS and Slot Times Defined*, IEEE 802.11 Working Group paper IEEE 802.11-95/81, May 1995.