IEEE 802.11 Throughput and Delay Analysis for mixed real time and normal data traffic

Matias Arenas P. and Walter Grote H.
Department of Electronic Engineering, Federico Santa Maria Technical U.
Avenida España 1680, Valparaíso, Chile

Abstract. IEEE 802.11 based network analysis has been largely focused on throughput performance. There has been a growing concern to provide quality of service (QoS) to this protocol suite, the result of which has been the considerable work towards the formulation of the IEEE 802.11e and IEEE 802.11n versions. One important aspect to consider is performance for real time applications like voice over IP (VoIP). In this paper we focus on performance issues of delay and throughput as a function of packet size, initial contention window size settings and the number of active terminals competing for access to the network, when some terminals provide VoIP services, while others transfer data. The simulation model developed using Network Simulator 2 (ns-2) is first validated comparing published results for throughput and delay, to then proceed to perform analysis on Ad hoc networks that will carry mixed VoIP and data traffic. We determine that tuning the initial and final contention window settings on the real time terminals does have a great impact on delay, throughput and packet loss in individual and on the network performance when in congestion.

1 Introduction

Wireless access technologies have experienced a tremendous growth in the last decade. End users have been attracted by various aspects of it, the most important being mobility. The IEEE 802.11 protocol provides wideband data services using small coverage cells, with distances ranging from 50m in in-building applications to about 300m in open spaces. These networks may work as pure random access networks using the DCF – Distributed Coordination Function – mode. Another possibility is to combine random access (DCF) and transmission scheduling (PCF:

Point Coordination Function) for delay sensitive information. Unfortunately, most implementations only consider DCF operation. Therefore it is interesting to analyze if delay sensitive information like VoIP can be delivered satisfactorily in networks that will supply data and voice transmission services using DCF. Random access can be used in peer to peer - also known as Ad Hoc – or centralized – also known as infrastructure – kind of networks, [1]. Considerable effort has been made to develop analytical and simulation models to establish throughput performance for random access networks to transport low-bandwidth, data application traffic, [2, 3, 4, 5, 6]. Today's requirements for wireless transmission for data applications requiring large bandwidth in conjunction with time-sensitive multimedia applications with quality of service (QoS) puts the focus on performance issues like delay, throughput and packet loss performance as a function of packet size, initial contention window size settings and the number of active terminals competing for access to the network, [7]. Reducing initial and final contention window settings, as well as the number of retransmissions at terminals running time sensitive applications will reduce delay and its jitter. However these adjustments come at a price, affecting overall delay and throughput, when in congestion. This publication aims to provide some further insight on these issues. Specifically we want to establish how to set the initial contention window size in terminals that will provide VoIP services in order to reduce delay in packet delivery, while not affecting global throughput and delay significantly.

Networks covering indoor spaces span short distances and problems like the hidden terminal and capture seldom occur. When the hidden terminal problem is present, a transmission or a collision of 2 or more transmissions may not be detected by some terminals that in turn may start to broadcast on their own, either creating a collision or contributing to an existing one [2]. Capture may occur when the received power from two terminals differ by a large amount due to the fact that one terminal experiences larger propagation losses than the other, [3]. In our analysis we consider that neither the hidden terminal nor the capture effect are present, and therefore consider only the basic access mechanism.

In [4], a simple but accurate, analytical model has been developed to compute the 802.11 Distributed Coordination Function (DCF) throughput in the assumption of finite number of terminals and ideal channel conditions. By means of the proposed model, an extensive throughput performance evaluation of both access mechanisms of the 802.11 protocol is provided. This analysis is a good starting point to develop an understanding of how the initial contention window (*CWmin*) size affects network performance. However it assumes that all terminals are configured the same way and it pays no attention to the effect of setting the maximum contention window (*CWmax*) size.

DCF uses a contention window (CW) to control the random access to the channel. Basically, it consists of a backoff counter that inhibits a terminal from immediate transmission, by delaying that instant to the moment the counter reaches zero, starting from an initial value that is being set upon arrival of the packet. The contention window is defined by two parameters: *CWmin*, *CWmax*. The random number used in the random backoff is initially a number between 0 and *CWmin*. If the initial backoff expires without successful broadcast of the packet, the terminal doubles the value random backoff window size ($CW = 2 \cdot CWmin$) and picks a new value at random between 0 and (*CW*-1). This doubling in size will continue with each additional retry until *CW* = *CWmax*. Once this value has been reached, further retries will be made picking a random number in the range [0, *CWmax*-1]. Retries continue until the maximum retries has been reached. This process of doubling the backoff window is often referred to as a binary exponential backoff. The influence of the backoff algorithm has been studied analytically and by simulation by [4, 5, 6, 7 and 8]. In [5, 6] Wu, et. al. enhance the analytical model developed by Bianchi, [4], limiting the maximum number of retries and define new metrics like goodput, fairness and average delay. Ziouva and Antonakopoulos obtain average delay measures, [7]. Xiao shows that a throughput upper limit and a delay lower limit exist since the overhead in the MAC magnifies itself when the data rate becomes higher, [8]. The lack of a built-in mechanism to provide quality of service with IEEE 802.11 based WLANs has triggered the work of a working group on a new standard, known as the IEEE 802.11e which introduces the so-called hybrid coordination function (HCF) for enhanced distributed channel access (EDCA) and HCF controlled channel access (HCCA). As a step in that direction, Cisco recommends to change settings of the contention windows at the Access Point (AP) thus being able to respond to class of service labels of appropriately tagged packets in infrastructure networks, [9]. Banchs and Vollero analyze the delay behaviour of the EDCA mechanism by varying the values of *CWmin, CWmax* and a parameter that determines how long a station has to wait to decrement the backoff counter after a successful transmission, AIFS, [10]. Wang et.al, analyze how to improve network performance, when VoIP traffic is considered, by changing some parameters at the AP of an infrastructure network, [11]. However, these studies consider that parameter settings of all stations as a whole, but do not consider the possibility of individual settings.

The difficulty associated to carry out a theoretical analysis when terminals possibly will be carrying traffic of different nature and may be configured with different parameter values invites to study network performance by means of a simulation model using Network Simulator 2 (ns-2). The aim of this publication is to analyze on how tuning the initial and final contention window settings on the terminals carrying real time traffic will affect delay and throughput in individual and global performance for Ad hoc networks that will carry mixed VoIP and data traffic under congestion.

2 Simulation Scenario

To establish a performance evaluation of the protocol with ns-2 simulator, we consider two kinds of terminals: some running real time applications with small sized packets scheduled for transmission (for example, VoIP, in which case a payload size of 55 bytes may be considered representative) and others transferring large files (for example, 1500 bytes, which is the maximum size of an Ethernet payload) with less stringent delay requirements (videostreaming, or data exchange). The hidden terminal and capture phenomena are avoided by placing terminals at equal distance (2m) from the center of a circle. All terminals transmit packets to one terminal that is placed in the center of the circle, which only receives. This arrangement is a simple means to keep track of all packets that have been received successfully. The network operates in saturation mode, that is, every transmitting terminal has always a packet ready for broadcast in its output buffer.

The ns-2 simulator is furthermore configured with following settings: *WirelessChannel*, *Two Ray Ground* as a radio propagation model, *Wireless physical* interface, *802.11 MAC*, *DropTail/PriQueue* queue management, the maximum queue length is 5 packets, *LL* link layer, omni-directional antennas, *DSDV* routing protocol. We have configured the slot time in 20[μs], SIFS time in 10[μs], *Preamble Length* in 144[bits], *PCLPHeaderLength* 48[bits], *Short Retry Limit* and *Long Retry Limit* are set to 7, unless stated otherwise. The basic rate and *PCLPDataRate* were set to 1[Mbps], *RTS threshold* to 3000 and packets were sent without ARP IP packet header. The physical space of the simulation is a circle with a radius of 500 meters. For each station, we run a CBR agent over UDP protocol. We set the data rate to saturate the network each simulation. We run the simulation for 200 [s] taking out the necessary information of the traces to remove warm up time data.

We setup the following experiments:

1. We set *CWmin* = 32, *CWmax* = 1024 at 6 of 7 stations. We set *CWmin* = 4, and *CWmax* = 4, 8, 16, 32, 64, 128, 256 and 512 in increasing values for each simulation at the trial station. Payload sizes are 1028 bytes for all packets. Transmission rate is 1 Mbps. The aim of this experiment is to determine the sensitivity of the network performance to the variation of the *CWmax* parameter at one station.
2. In the second experiment we try to evaluate the effect of having an increasing number of stations that reduce *CWmax*. We initially set *CWmin* = 32 and *CWmax* = 1024 at all stations, to reduce it to *CWmin* = 4 and *CWmax* = 64 in an increasing number of stations. Since 7 stations are involved, 8 simulations are possible and we number them as we increase the number of stations using a reduced *CWmin, CWmax* setting. Payload packet size is 1028 bytes. Transmission rate is 1 Mbps.
3. The third experiment is aimed to determine the effect of packet and contention window size in a network were all but one station of the network of 8 stations use the default settings of *CWmin* = 31 and *CWmax* = 1024, except the remaining station, that uses *CWmin* = 4 and *CWmax* = 8. The number of retries at all stations is set to 7. Transmission rate is 1 Mbps.
4. Experiment 4 considers two terminals that only transmit payload packets of 55 bytes. We set the default values *CWmin* = 32 and *CWmax* = 1024 for the backoff algorithm of these terminals. Data terminals are added in pairs to the network, transmitting 1500 bytes payload packets with the same default settings of the backoff algorithm. Transmission rate is 1 Mbps. We look for average delay, delay jitter and throughput globally and for each focus group (data and voice). We also look for the standard deviation of these performance measures. The aim of this experiment is to establish how the presence of an increasing number of data terminals affects voice connections when default settings are being used in terminals.
5. Same configurations as in experiment 4, but at 11 Mbps rate. With higher transmission rates, it is to be expected that more bits per second will reach the destination successfully, on the average. However, since control information is transmitted at a lower data rate than user data, more time will

be spent in transmitting control information. We want to make an assessment of how this affects overall performance.

6. In experiment 6 two terminals transmit only VoIP size packets (55 bytes) with the contention window set to *CWmin* = 4, *CWmax* = 16. We add 2 terminals at a time, with default settings of the backoff algorithm parameters (*CWmin* = 32, *CWmax* = 1024), transmitting 1500 bytes packets in pairs. We look for average delay, jitter and throughput globally and for the two focus groups on the average and establish its standard deviation. Transmission rate is 1 Mbps. The aim of this experiment is to see the advantage that can be obtained by reducing the contention window settings of the backoff algorithm to reduce delay of voice traffic, as compared to experiment 4.
7. Same configuration as in experiment 6, but transmissions are at at 11 Mbps rate, so as to compare the effect of the contention window size reduction for voice traffic.

These experiments will provide insight of network performance experiencing congestion.

3 Experiment Outcomes

We here describe the outcomes of the experiments outlined in the previous section. To get these results we program a trace filter in C language to eliminate warm up time and system messages from the *ns-2* simulation. Thereafter we identify packets transmitted by each station and proceed to find the time elapsed from the moment a packet is ready for transmission and acknowledgement reception. We then establish average value, 95% confidence intervals of the average value and standard deviation using Matlab.

Table 1 shows the results of experiment 1, with the first column showing the values of *CWmax* of the single station that varies its backoff algorithm configuration. The second column shows the lower limit of the 95% confidence interval of the average delay, while the third column shows the average delay, the fourth column the upper limit of the 95% confidence interval of the average delay and the fifth column, the standard deviation of that delay for the single station. Columns 6, 7, 8 and 9 are the set of values for the remaining seven stations of the network. Columns 10, 11, 12 and 13 reflect the values of the average delay and standard deviation of the entire network (all 8 stations). Payload sizes of packets are 1028 bytes for all terminals.

From the data displayed in table 1 one may easily conclude that while overall network delay remains approximately the same, the use of low values of *CWmax* in one station reduces considerably its own delay to less than 59% of the delay experienced by the remaining stations, on the average, while *CWmax* of that station is kept below 32. If *CWmax=64* for the single station, delays are still 69% of the average delay experienced by the remaining stations. For values larger than *CWmax* = 64 on the single station, the delay is basically the same for all stations and has no effect on overall performance. This result coincides with a recommendation issued by CISCO for Access Point settings in infrastructure networks, which states that one should avoid setting *CWmax* at delay sensitive stations below *CWmin* of the rest of the stations, so as not to affect overall network performance, [9]. However, if the aim of the adjustment is to reduce time response at connections sensitive to delay, this result clearly indicates the convenience of setting *CWmin* and *CWmax* of that connection at lower values than the rest of the network settings for that purpose.

Table 1. Network performance sensitivity to *CWmax* variations on one station.

1 Mbps	Single station, CWmin = 4 Delay [s]				Remaining stations, CWmin = 32, CWmax = 1024 Delay [s]				Network values Delay [s]			
CWmax of Single Station	Low 95% CI Value	Average	Upper 95% CI Value	Std. Dev.	Low 95% CI Value	Average	Upper 95% CI Value	Std. Dev.	Low 95% CI Value	Average	Upper 95% CI Value	Std. Dev.
4	0.0438	**0.0447**	0.0456	0.0041	0.0808	**0.0825**	0.0841	0.0135	0.0716	**0.0734**	0.0752	0.0236
8	0.0454	**0.0464**	0.0473	0.0043	0.0796	**0.0815**	0.0834	0.0150	0.0716	**0.0734**	0.0752	0.0238
16	0.0482	**0.0494**	0.0505	0.0047	0.0815	**0.0838**	0.0861	0.0187	0.0742	**0.0764**	0.0785	0.0277
32	0.0532	**0.0545**	0.0559	0.0055	0.0778	**0.0797**	0.0815	0.0154	0.0733	**0.0750**	0.0767	0.0216
64	0.0665	**0.0693**	0.0722	0.0101	0.0747	**0.0762**	0.0778	0.0133	0.0739	**0.0754**	0.0768	0.0188
128	0.0677	**0.0712**	0.0748	0.0121	0.0741	**0.0758**	0.0775	0.0146	0.0736	**0.0751**	0.0766	0.0195
256	0.0734	**0.0775**	0.0816	0.0134	0.0730	**0.0746**	0.0762	0.0135	0.0734	**0.0748**	0.0762	0.0183
512	0.0726	**0.0774**	0.0822	0.0156	0.0744	**0.0763**	0.0781	0.0160	0.0746	**0.0762**	0.0779	0.0218
1024	0.0684	**0.0716**	0.0749	0.0108	0.0734	**0.0751**	0.0767	0.0141	0.0733	**0.0747**	0.0761	0.0186

In Table 2 we show the results of experiment 2. The first column the number of stations configured with *CWmin* = 4 and *CWmax* = 64, while the rest of the stations use the default values of *CWmin* = 32 and *CWmax* = 1024. That is, at row 0 the simulation is run with all stations using the default settings for the backoff algorithm, at row 1, one station is set to *CWmin* = 4 and *CWmax* = 64, while the rest maintains the default settings, and so on. The remaining columns have a similar meaning as it was explained in Table 1.

Table 2. Delays when an increasing number of stations reduce *CWmin* and *CWmax.*

1Mbps Number of Stations with reduced CW	Stations with CWmin = 4, CWmax = 64 Delay [s]				Remaining stations, with default CW,CWmin = 32, CWmax = 1024 Delay [s]				Network values Delay [s]			
	Low 95% CI Value	Average	Upper 95% CI Value	Std. Dev.	Low 95% CI Value	Average	Upper 95% CI Value	Std. Dev.	Low 95% CI Value	Average	Upper 95% CI Value	Std. Dev.
0	0.0351	**0.0369**	0.0389	0.0105	0.0351	**0.0360**	0.0369	0.0121	0.0354	**0.0362**	0.0371	0.0124
1	0.0223	**0.0228**	0.0232	0.0034	0.0395	**0.0405**	0.0415	0.0130	0.0354	**0.0363**	0.0373	0.0139
2	0.0244	**0.0249**	0.0254	0.0036	0.0383	**0.0397**	0.0411	0.0178	0.0352	**0.0364**	0.0375	0.0171
3	0.0267	**0.0272**	0.0278	0.0038	0.0385	**0.0401**	0.0417	0.0204	0.0362	**0.0375**	0.0388	0.0190
4	0.0283	**0.0290**	0.0297	0.0044	0.0378	**0.0396**	0.0414	0.0232	0.0360	**0.0375**	0.0390	0.0218
5	0.0328	**0.0336**	0.0344	0.0047	0.0391	**0.0411**	0.0431	0.0261	0.0381	**0.0397**	0.0414	0.0239
6	0.0377	**0.0385**	0.0394	0.0047	0.0402	**0.0416**	0.0430	0.0185	0.0398	**0.0410**	0.0422	0.0166
7	0.0439	**0.0449**	0.0458	0.0050	0.0436	**0.0441**	0.0445	0.0057	0.0438	**0.0442**	0.0446	0.0056

Reading the data of table 2 shows that when less than 4 stations switch to lower values of *CWmin* and *CWmax*, these stations will experiment delay reduction of 73% as compared to the remaining stations running with the default settings of the backoff algorithm. On the other extreme, if 7 stations switch to the lower values of *CWmin* and *CWmax*, the average delay for these stations is worse than for those that use the default values (compare row 0 to row 7) and the overall network delay has increased due to the fact that the number of collisions has increased. This tells us that only a fraction of all stations of a network should be privileged, if deemed necessary, with a reduction of their *CWmin* and *CWmax* values. These stations should be running real time applications. Thus, one may conclude that in a network running VoIP and data connections it seems to be a safe practice to reduce the contention window size of the backoff algorithm of terminals dealing with voice traffic on a regular basis.

Figure 3 reflects the effect of packet and contention window size in a network were all but one station of the network of 8 stations use the default settings of *CWmin* = 31 and *CWmax* = 1024, except the remaining station, that uses *CWmin* = 4 and *CWmax* = 64. The number of retries at all stations is set to 7. Transmission rate is 1 Mbps.

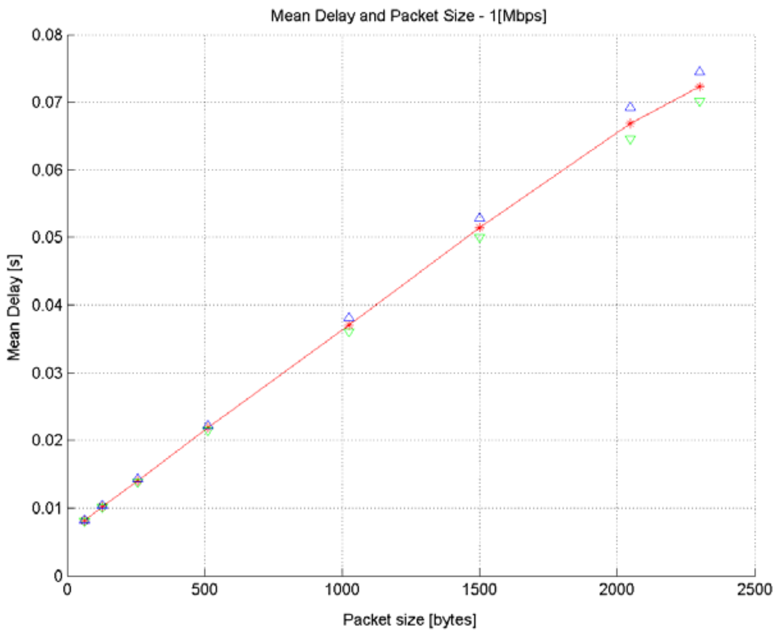

Fig. 3. Effect of backoff algorithm configuration and packet size in network.

From figure 3 one may conclude that at a given transmission rate, if all stations transmit packets of equal size, the delay that each successful transmission experiences is proportional to the packet size, a fact that is intuitively perceived. The fact of having one terminal configured with a lower value of the initial contention window does not seem to affect the overall performance of the network. Therefore, if a terminal is configured to have a lower value of the initial contention window size to reduce its delay when transmitting voice packets and then uses these settings to transmit packets of different sizes, as any other terminal will do, the effects on overall performance are negligible.

Experiment 4, 5, 6 and 7 are an attempt to establish the viability of having a terminal carrying delay sensitive data (VoIP) in a network that otherwise provides a wireless service to data transfers. The voice terminals have been configured with either *CWmin* = 32 and *CWmax* =1024 or *CWmin* = 4 and *CWmax* = 16, while data terminals have been set with *CWmin* = 32 and *CWmax* = 1024, according to the findings of experiment 1. Figure 4 shows the results for a network operating at 1 Mbps. The number of voice stations (VS) and stations sending data packets (DS) are shown on the horizontal axis. Average values of the 95% upper and lower confidence intervals are indicated. Figure 5 exhibits a similar result for network operating at 11 Mbps.

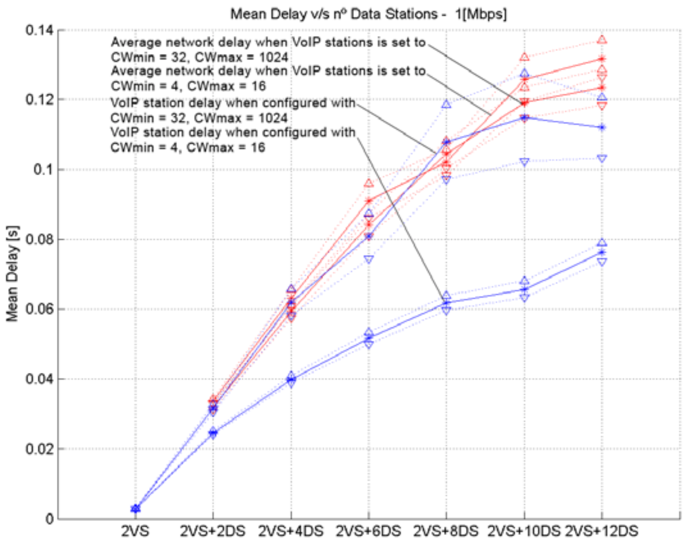

Fig. 4. Delay performance of VoIP and data terminals operating at 1 Mbps.

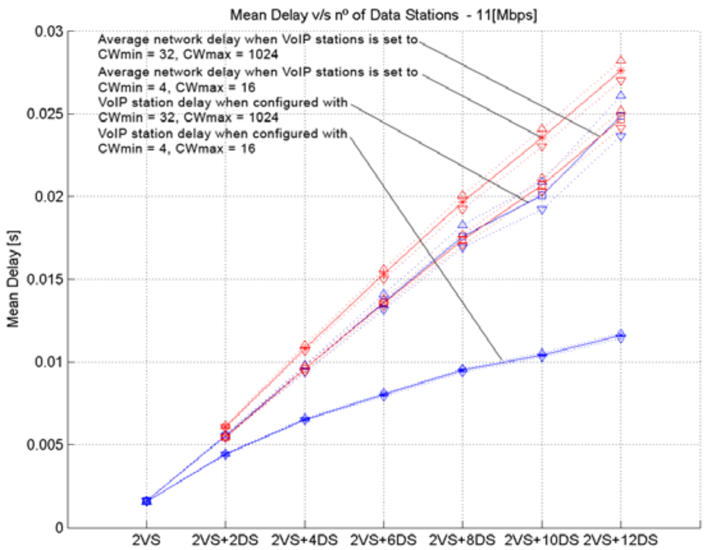

Fig. 5. Delay performance of VoIP and data terminals operating at 11 Mbps.

From figure 4 and 5 it is simple to see that there is a clear advantage for VoIP connections in terms of delay if the initial and maximum contention window values are set to *CWmin* = 4 and *CWmax* = 16, while the rest of the stations maintain the default settings (*CWmin* = 32 and *CWmax* = 1028). One advantage of assigning access privileges to delay sensitive terminals is that not only delay is reduced, but so is its jitter, when the network consists of many terminals. Another advantage is that even though VoIP terminals have improved their performance by having been granted access privileges, the affect on delay and delay jitter on data terminals is almost not perceived. However, these settings do affect overall network performance as can be seen in figure 6. Overall throughput deteriorates due to the increase in collisions as a result of the reduction of the window sizes of the backoff algorithm of the VoIP terminals.

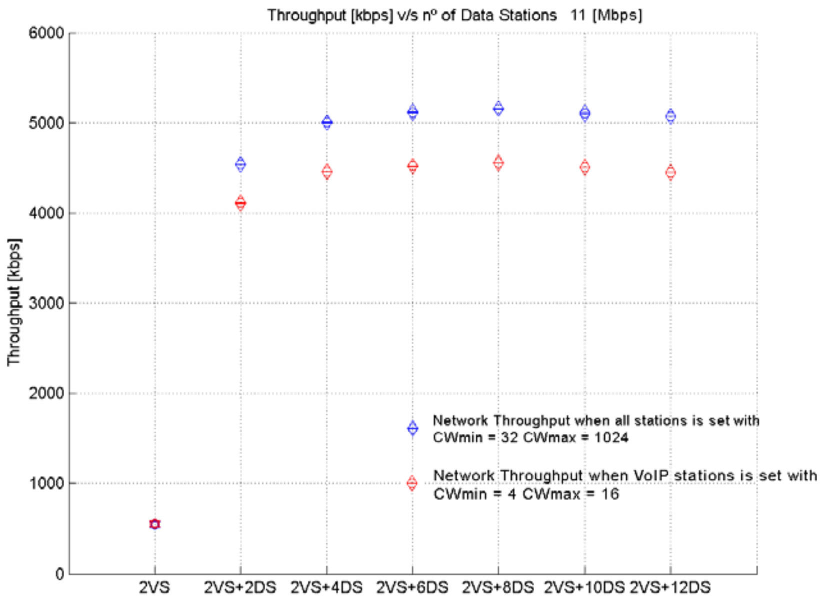

Fig. 6. Throughput as a function of network load and contention window settings.

In figure 6 the upper throughput points are due to a network that has all stations working with the default contention window configuration (*CWmin* = 32, *CWmax* = 1024). The lower values belong to the network where data stations (DS) use the default settings, while the 2 voice stations (VS) have their values set to *CWmin* = 4 and *CWmax* = 16. Clearly, a 10% deterioration is observed due to the overhead of collisions. However, the advantage observed is a 50% reduction of the delay of the time sensitive application running on the voice stations (see figure 5).

Conclusions

We have conducted a set of simulation experiments for wireless Ad Hoc wireless networks using the IEEE 802.11 protocol to be able to establish in which way time sensitive applications may benefit from reducing the contention window settings (*CWmin* and *CWmax*) of the backoff algorithm, thus reducing their delay without affecting the average network delay that much, as long as only a few stations of the network take advantage of this possibility. Throughput degradation may be acceptable under these circumstances as a necessary tradeoff due to the overhead caused by an increase of collisions and increased overhead due to the smaller packet sizes of the voice connection. It is in our best knowledge that no publication has reported this effect so far.

Acknowledgements

This publication was made possible with partial funding from project PBCT ACT-11-04 and project UTFSM 23.05.21.

References

1. ANSI/IEEE Std 802.11, Part 11: Wireless LAN Medium Access, Control (MAC) and Physical Layer (PHY) Specifications, 1999 Edition (R2003)
2. S.Khurana, A.Kahol and K.Srimani, Performance evaluation of distributed co-ordination function for IEEE 802.11 wireless LAN protocol in the presence of mobile and hidden terminals, 7th Intl Symp. On Model., Anal. And Simul. Of Comp. and Telecom. Sys. (MASCOT99), October 1999.
3. Z.Hadzi-Velkov; B. Spasenovski; On the impact of antenna diversity in IEEE 802.11b DCF with capture, TELSIKS 2003. Conference, Vol. 1, 1-3 Oct. 2003 Page(s):93 – 96.
4. G. Bianchi, Performance Analysis of the IEEE 802.11 Distributed Coordination Function, IEEE Journal on Selected Areas on Communications, Vol. 18, N°. 3, March 2000, pp. 535-548.
5. H. Wu, Y. Peng, K. Long y S. Cheng. ``A Simple Model of IEEE 802.11 Wireless LAN". Proc. IEEE Intertational Conferences on Info-Tech and Info-net (ICII), Beijing, vol 2, 514-519, Oct. 2001.
6. H. Wu, Y. Peng, K. Long, S. Cheng, J. Ma, "Performance of Reliable Transport Protocol over IEEE 802.11 Wireless LAN Analysis and Enhancement", IEEE INFOCOM Feb. 2002.
7. E. Ziouva y T. Antonakopoulus, "The Effect of finite population on IEEE802.11 Wireless LANs Thoughput/Delay Performance", Proc 11th IEEE MELECON 2002, pp. 95-99, May 7-9 .
8. Y. Xiao, ``A Simple and Effective Priority Scheme for IEEE 802.11". IEEE Comm. Letters, 7(2): 70-72 Feb 2003

9. http://www.cisco.com/en/US/products/hw/wireless/ps430/prod_technical_reference09186a0080144498.html
10. A. Banchs, L. Vollero, A delay model for IEEE 802.11e EDCA, IEEE Communications Letters, Vol. 9, N° 6, Jun 2005, pp. 508- 510
11. W. Wang, S. Chang, L. Li, Solutions to performance problems in VoIP over a 802.11 wireless LAN, IEEE Trans. on Vehic. Tech., Jan. 2005, Vol. 54, N°: 1, pp. 366- 384