802.11 Buffers: When Bigger Is Not Better?[⋆]

David Malone[1], Hanghang Qi[1], Dmitri Botvich[2], and Paul Patras[1]

[1] Hamilton Institute, National University of Ireland Maynooth
[2] TSSG, Waterford Institute of Technology, Ireland

Abstract. While there have been considerable advances in the modelling of 802.11's MAC layer in recent years, 802.11 with finite buffer space is considered difficult to analyse. In this paper, we study the impact of finite buffers' effect on the 802.11 performance, in view of the requirements of interactive applications sensitive to delay and packet loss. Using both state-of-the art and simplified queueing models, we identify a surprising result. Specifically, we find that increased buffering throughout an 802.11 network will not only incur delay, but may actually increase the packet loss experienced by stations. By means of numerical analysis and simulations we show that this non-monotonic behaviour arises because of the contention-based nature of the medium access protocol, whose performance is closely related to the traffic load and the buffer size. Finally, we discuss on protocol and buffer tuning towards eliminating such undesirable effect.

1 Introduction

The IEEE 802.11 protocol has grown to be the de facto standard for wireless LANs since it was developed and released in the 1990s. While the 802.11 specification includes both centralised and decentralised MAC mechanisms, the distributed coordination function (DCF), a random access scheme based on carrier sense multiple access with collision avoidance (CSMA/CA) through binary exponential backoff (BEB), is the scheme widely used in current devices. Modeling the 802.11 DCF performance and subsequent protocol optimisations have advanced considerably over the last 10–15 years. In particular, Bianchi's technique [3], has permitted the performance analysis of a number of protocol variations. This approach models the MAC protocol states, in particular the backoff counter and backoff stage, as a Markov chain. Precisely, the analysis leads to a Markov chain with $O(1000)$ states and by finding its stationary distribution, one can predict the throughput of the network.

For tractability, Bianchi studied an 802.11 network where stations always had packets to transmit (i.e. saturation conditions). In this situation, queueing dynamics can be ignored, which simplifies the model but also does not capture

accurately the protocol behaviour under realistic traffic. Building on this, other authors considered non-saturated networks with small buffers (e.g. [10]) or with infinite buffers (e.g. [12,6]). In the former case, queue state is simple and does not significantly complicate the model. In the latter case, if the analysis of the queue and the MAC are decoupled by assuming they operate probabilistically independently, then both the queueing and MAC parts of the model remain relatively simple, and so the analysis remains tractable.

Performance of 802.11 with finite buffers is commonly considered to be difficult to analyse. If one couples the MAC and the queue, the result is a larger Markov chain whose stationary distribution is hard to calculate explicitly. The usual formulation results in a Markov chain with $O(1000K)$ states, where K is the number of packets that can be buffered. This has resulted in a number of models which are tractable as long as buffering is limited to a modest number of packets. Alternatively, if one decouples the MAC and the queue, the result is a reduction in accuracy of the predictions [7]. Liu et al. [9] introduced a model that retains much of the coupling between the MAC and the queue, while they also provide numerical techniques for efficiently solving for the Markov chain's stationary distribution.

In this paper, we present a study of the 802.11 behaviour with finite-load finite-buffer, by building upon the state-of-the-art Markov chain model of Liu et al. Specifically, we are interested in the case where small buffers may offer a performance advantage to delay-sensitive traffic, such as online games and conversational video/voice. We show that increasing buffer sizes may actually *increase* the packet loss rate, while also increasing delays. To demonstrate that this is not an artefact of the model Liu et al. proposed, we show that the same effect is present also in a much simpler model that combines queueing and contention. Using numerical analysis and simulations, we illustrate that this effect arises due to the contention-based access to the wireless medium. Finally, based on these findings, we discuss on buffer tuning and MAC configuration towards improving the performance of interactive applications.

2   802.11 Modelling with Finite Buffers

To study the impact of the buffer size of the 802.11 performance, we employ two analytical models: the model introduced by Liu et al. [9], which integrates the queueing and MAC operation, and a simplified $M/M/1/K$ station model that we solve.

2.1   Markov Chain Model of Liu et al.

The details of the model proposed by Liu et al. can be found in [9] with an extensive validation. Here we only summarise the key results that we use for our analysis. The Markov chain therein has 3-dimensional state space consisting of 802.11's backoff counter, backoff stage and the current queue length. As in Bianchi's model, the chain is not real-time, but evolves upon counter decrements at the MAC layer. Certain approximations are made for tractability, e.g.

the number of arrivals in a state change is limited to one packet. The authors give a method for solving for the stationary distribution of the Markov chain using linear algebra techniques, which involves considerably lower complexity as compared to brute-force methods.

Of particular interest to our study is the expression of the total delay (due to both queueing and contention), which if we consider frames do not exceed the maximum retry limit, can be computed as [9]:

$$D_t = \Delta \left(1 + \sum_{h=0}^{K} h P[Q = h] \right),$$

where Δ denotes the MAC delay, i.e. the average time elapsed since a packet reaches head-of-the-line until its successful transmission, $P[Q = h]$ defines the probability of the queue being h at any given time, and K denotes the maximum queue length. Further, under the same assumptions the packet loss rate due to buffer being full (i.e. blocking probability) is given by:

$$P_B = P[Q = K].$$

To study the effect of the buffer size on the delay and packet loss performance, we have independently implemented this model. Comparing the output with the figures reported in [9], we confirm the accuracy of our implementation.

2.2 M/M/1/K Simplified Analysis

For comparison, we also consider a simplified model, where we replace the 802.11 DCF MAC with a simple slotted Aloha-style network. We still view each station as a queue of size K, but use an $M/M/1/K$ approximation to model its behaviour. The $M/M/1/K$ approximation assumes that queue inter-arrival times and service times are independent and exponentially distributed, which allows us to explicitly calculate many queue statistics and will prove useful in further validating the findings we obtain by employing the model of Liu et al. Thus we will solve for the probability of a successful transmission in terms of the Aloha transmission probability, the size of the network, the traffic arrival rate and buffer size. Based on this, we derive the blocking probability and total delay, to examine the buffer size impact on the performance.

Suppose we have n queues feeding a slotted Aloha style network where each queue transmits in a slot with probability τ_0 when it has a packet. We will assume the event that a queue transmits in each slot is independent and identically distributed. Let P_{ne} be the probability that a queue is not empty. Let $\tau = P_{ne}\tau_0$ be the probability that each node transmits. The service rate for a queue is the probability that a queue transmits with no collision, or

$$\mu = C(\tau) = \tau_0 (1 - \tau)^{n-1},$$

and service times will actually be geometrically distributed.

Now suppose arrivals to the queue occur at a fixed rate of λ packets per slot. Consequently, we have a traffic intensity of $\rho = \lambda/\mu$. Then, approximating the queue by an $M/M/1/K$ model [1], we can compute the queue non-empty probability as

$$P_{ne} = \frac{\rho - \rho^{k+1}}{1 - \rho^{k+1}}.$$

Thus we can regard $P_{ne} = Q(\mu/\lambda)$, remembering that λ is fixed. So, the operating point of the network is given by a fixed point characterised by:

$$\mu = C(\tau) = C(P_{ne}\tau_0) = C(Q(\mu/\lambda)\tau_0).$$

Since τ_0 and λ are given parameters, we can solve this numerically to find μ and then substitute to find P_{ne}.
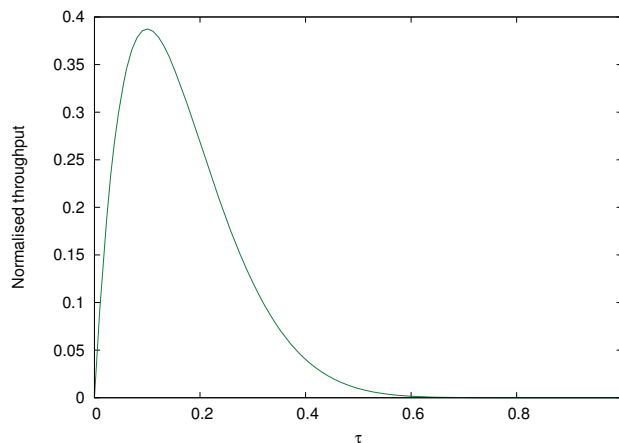
Fig. 1. Throughput for a slotted Aloha network with $n = 10$ stations

Note that for standard slotted Aloha networks, it is well-known [2] that the value of τ which gives optimal throughput is $1/n$. This is illustrated in Figure 1 for $n = 10$ stations, where we can see a clear peak in the total throughput of the network at $\tau = 0.1$. Below this peak, stations do not transmit sufficiently often, leaving too many slots idle, whereas the network experiences an increased number of collisions beyond the peak, which results in reduced throughput.

Having computed μ and knowing λ, we can obtain the packet loss probability due to buffer being full, as follows

$$P_B = \frac{1 - \rho}{1 - \rho^{K+1}}\rho^K.$$

To derive the expression of the total delay a packet undergoes from the time of arrival to the queue until successful transmission, we first compute the average queue length,

$$\bar{N} = \frac{1-\rho}{1-\rho^{K+1}} \sum_{k=0}^{K} k\rho^k$$

and then apply Little's law [8], which yields

$$D_t = \frac{\bar{N}}{\lambda}.$$

3 Performance Analysis

In what follows we analyse the impact of the buffer size on the packet loss and delay performance of the networks, using the models detailed in Sec. 2. Consider a symmetric network, in the sense that all stations are have the same buffer size and operate under identical traffic load. Consequently, we expect each station to experience the same mean throughput, queueing delay and packet loss probability. Unless otherwise stated, we assume a network with $n = 10$ nodes that transmit frames with 500-Byte payload and employ the MAC and PHY layer parameters of IEEE 802.11b as detailed in Table 1.

3.1 Numerical Results with the Model of Liu et al.

First we investigate the packet loss rate and total delay predicted by the 3-D Markov chain model of Liu et al. [9], as the buffer size is varied. Initially we examine the performance under a light load regime, whereby the total offered load is 60% of the network's idealised total capacity, i.e. the number of packets that could be transmitted if the medium was occupied by back-to-back successful transmissions, spaced as closely as permitted by the protocol without backoff, which also the convention used in [9]. Note, however, that the practical capacity is somewhat lower due collisions and backoff procedure.

Figure 2 shows the packet loss rate and delay for this scenario. Since the network is lightly loaded, we observe a familiar pattern, i.e. the delay experienced by each packet is effectively constant. Also, as we increase the available buffer space, the packet loss rate decreases since it is less likely the queue will be full upon the arrival of a new packet.

Table 1. 802.11b MAC parameters

Basic rate	1 Mb/s	$CW_{min} = W$	32
Data rate	11 Mb/s	$m\ (CW_{max} = 2^m CW_{min})$	5
Preamble	144 bits	SIFS	10 μs
Headers	40 bytes	DIFS	50 μs
Payload $E[P]$	500 bytes	Slot time σ	20 μs
ACK	14 bytes	Propagation delay	1 μs
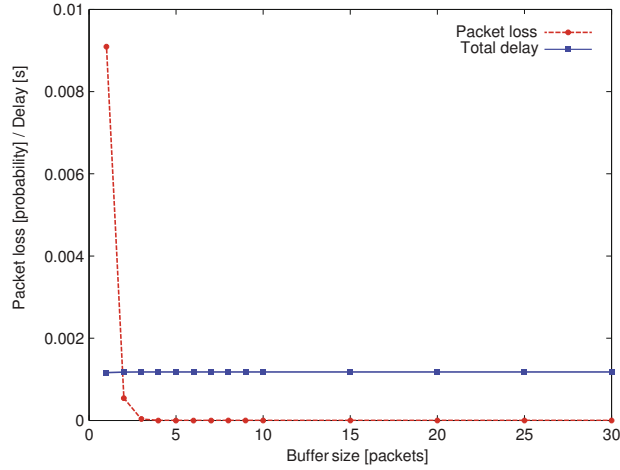
Fig. 2. Delay and packet loss as buffer size is varied. 60% total offered load.
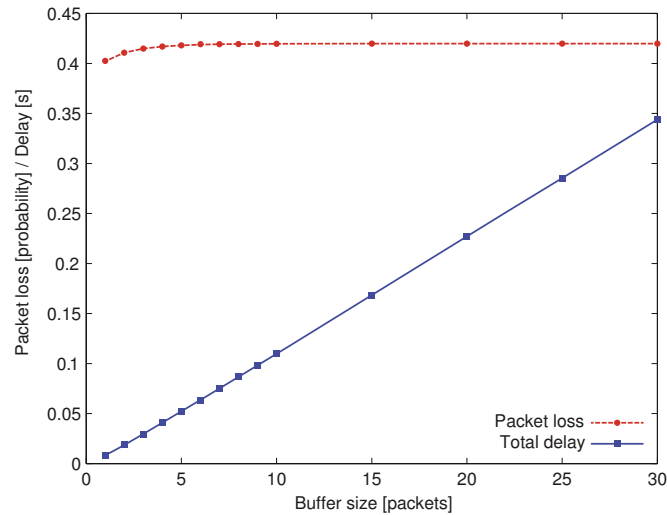
Fig. 3. Delay and loss as buffer size is varied. 140% total offered load.

Let us now examine the case where the network is heavily loaded, the load exceeding the network capacity. In this scenario, buffers become immediately full and the excess load is dropped as the queues cannot keep up with new arrivals. Figure 3 illustrates the performance in such circumstances. In this example, we consider an offered load of 140% of the total capacity, which results in slightly more than 40% of the packets being discarded. Also, as the queues remain full

the total delay is dominated by the queueing delay, which increases linearly with the buffer size.

What is more interesting to observe is the effect of the buffer size on the performance of the network under moderate-to-high loads. For this purpose, we analyse the packet loss rate and delay in a scenario where the total load is 85% of the network capacity. This results are depicted in Figure 4. As expected, since the traffic volume is significant (however, not overloading the network) we see the delay increasing as we increase the buffer size. On the other hand, the packet loss rate initially decreases as we increase the buffer size, but we find that beyond approximately 5 packets of buffering, surprisingly the packet loss rate actually increases as the queue expands.

We can gain some insight into why increasing buffering actually increases losses by considering the same network with $n = 10$ stations, where we fix the queue length of $n - 1$ of them to $K = 5$ and vary the buffer size of one other station. The resulting packet loss rates in this scenario are shown in Figure 5. We see that as we increase the buffer size of one station, the packet loss rate it experiences decreases, while the packet loss rate of the other stations in the network increases. This suggests that, as we increase contention for the wireless medium, by reducing the packet loss of one station, we are decreasing the time available to other stations through a mix of collisions and successful transmissions.

In the following subsection, we demonstrate this effect more clearly by using the simplified $M/M/1/K$ model introduced in Sec. 2.2.
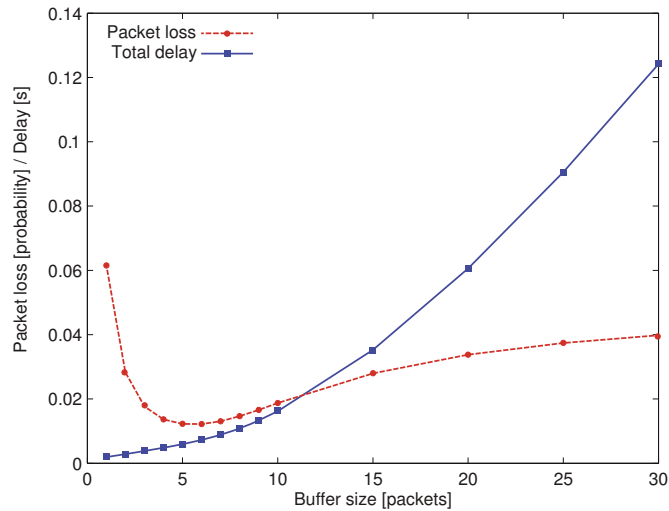

Fig. 4. Delay and packet loss as buffer size is varied. 85% total offered load.
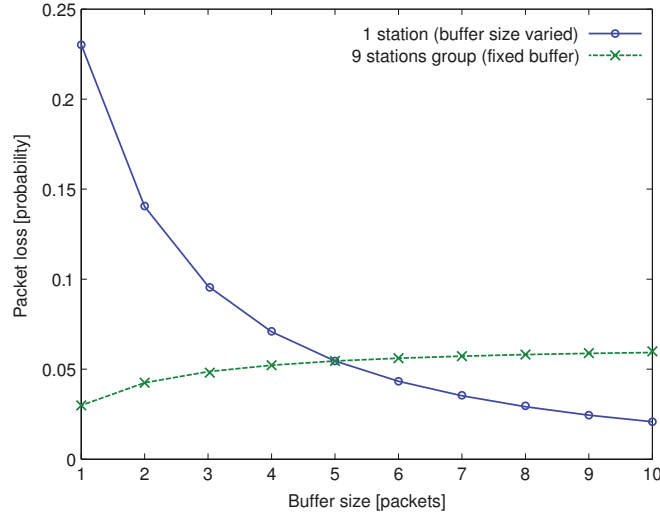
Fig. 5. Packet loss behaviour as buffer size is varied only for one station

3.2   Numerical Results with the M/M/1/K Model

To further illustrate the impact of the buffer size on the packet loss rate at moderate-to-high traffic loads, and confirm that the non-monotonic behaviour observed in our numerical analysis based on the model of Liu et al. is not an artefact of the model, here we consider an equivalent network with $n = 10$ active stations running slotted Aloha and study the packet loss as predicted by our $M/M/1/K$ model.

Although the network parameters are not directly comparable, we consider stations with a maximum transmission probability of $\tau_0 = 0.15$, which is larger than the optimal transmission probability of $1/10$. In this scenario, provided the offered load is large enough, by varying the buffer size we can move from a situation where the actual transmission probability is below the optimal value, to a situation where it exceeds the optimal value. For an offered load of $\lambda = 0.045$ packets per slot per station, the resulting packet loss is shown in Figure 6, where we identify a pattern very similar to the packet loss observed for the 802.11 case depicted in Figure 4.

To add more perspective, we also calculate the transmission probabilities of the stations and compare them to the optimal value of $1/10$. The results are shown in Figure 7. We see that as we increase the buffer size K, the actual transmission probability τ increases, indicating that there are packets available for transmission at the head of the queue more often. However, for the parameters shown, once $K > 4$ the value of τ exceeds the optimal value. As we saw in Figure 6, this results in increased packet loss, since the number of collisions caused by simultaneous transmissions outweighs the extra transmission attempts.
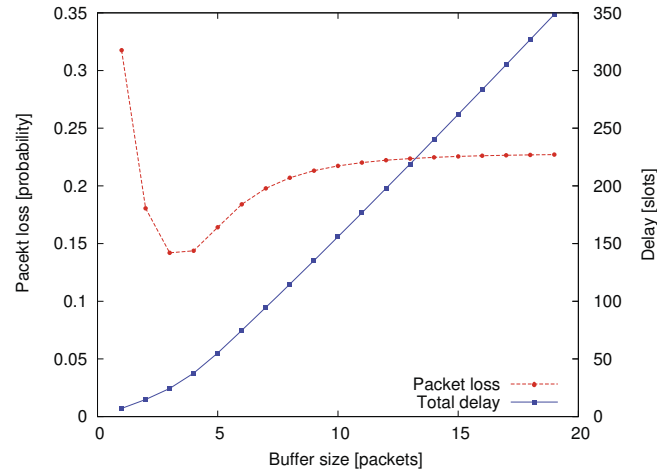
Fig. 6. Packet loss & delay behaviour as buffer size is varied. ALOHA network ($n = 10$).
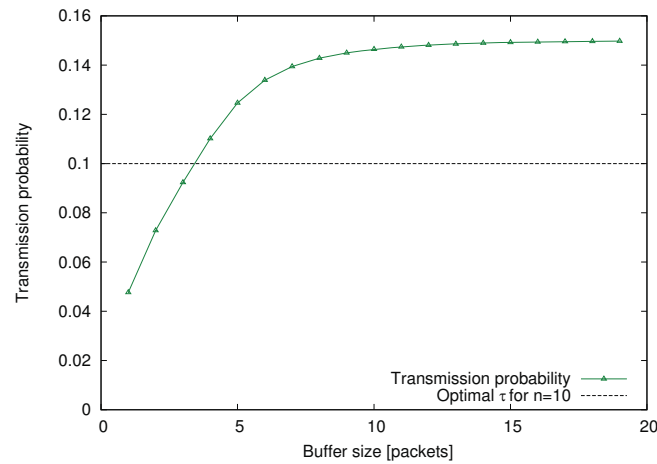
Fig. 7. Transmission probability as buffer size is varied

Following these observations, we believe that the behaviour shown by the 802.11 system in Figure 4 can be understood in the same way. This finding is in fact consistent with earlier work that demonstrates the throughput of an 802.11 network can decrease as the offered load increases even when buffering is fixed [10], as our results indicate that increased buffering may also increase the effective transmission probability for 802.11 beyond its optimal value, and consequently decrease network throughput.
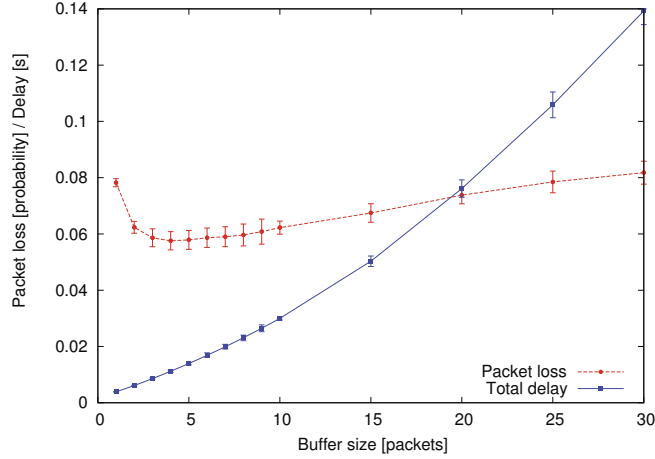
Fig. 8. Performance evaluation of a simulated network with real-time traffic
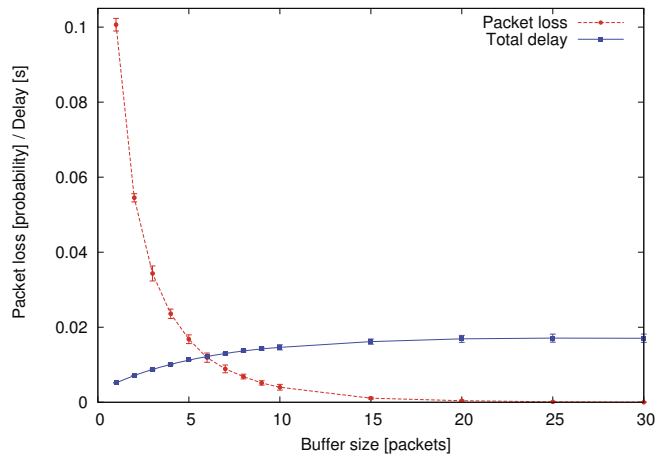

Fig. 9. Performance of an optimally configured network with real-time traffic

3.3   Simulation Results

To conclude our performance analysis, we consider a realistic scenario that is likely to be impacted by the identified non-monotonic behaviour of the packet loss when the buffer size is varied. Specifically, our focus is on interactive applications such as online games, video conferencing or voice over WLAN. Therefore, we run simulation experiments with a 10-node network, where each station runs a delay sensitive application that generates 425 kb/s of traffic, mapped over the voice queue at the MAC layer. For this purpose, we use an event-based simulator that we implemented in C/C++, which follows the 802.11 protocol rules and timing, and consider synthetic constant bit-rate traffic sources, generating 500-byte packets. Given the MAC layer configuration for the access category

employed ($CW_{min} = 8$, $CW_{max} = 16$) and the traffic volume, the network will be operating in a medium-to-high load regime. After a 30 s warm-up period, we measure the average packet loss rate and delay over a duration of 5 min. We repeat the simulations 10 times and plot the average and 95% confidence intervals for the two metrics in Figure 8. We identify the same trend as observed in the previous numerical analysis, which strengthens our belief that appropriate MAC and buffer tuning is necessary to optimise the performance of such applications.

4 Discussion

We have seen that due to the contention-based nature of the 802.11 protocol, packet loss may exhibit a non-monotonic behaviour when the buffer size is increased, as the transmission probability becomes larger than the optimal value. In addition, the total delay increases substantially, which may have a severe impact on interactive applications. On the other hand, such applications are less susceptible to packet loss [5]. In these circumstances it is important to appropriately tune the MAC parameters and buffer size of the stations, to meet latency requirements.

By setting the maximum transmission probability τ_0 (corresponding to saturation) of the nodes to a fixed value, one can ensure that the actual transmission probability will not increase beyond this point irrespective of the offered load. Further, if this is set to the optimal value, delay can be optimised, while the non-monotonic behaviour of the loss identified previously can be eliminated. Thus the total delay can be bounded by appropriate tuning of the buffer size.

Setting τ_0 to a fixed value, is easily achievable by configuring nodes with $CW = CW_{min} = CW_{max}$, i.e. $\tau_0 = 2/(CW + 1)$. For optimality, this can be performed either in a centralised way from the AP, which holds information about the network size (see [3] for a detailed discussion), or in a distributed manner, by only observing the channel conditions and dynamically tuning CW, as in e.g. [11]. With this setting, as long as the offered load does not exceed the service rate, a small buffer would suffice, while as stations become overloaded, delay is directly proportional to the buffer size. To demonstrate this, consider again the earlier simulation experiment, with $n = 10$ nodes sending each 425 kb/s to the AP, but this time contending with the optimal CW configuration. As shown in Figure 9, packet loss drops to zero as the buffer size exceeds 10 packets, while the delay remains bounded below 20 ms, which is perfectly suitable for the real-time traffic considered [4].

5 Conclusions

In this paper we have demonstrated that increasing the buffer size in an 802.11 network may actually increase packet loss rates. We demonstrate this result using a state-of-the-art Markov chain model of buffered 802.11 and a simpler model of a slotted Aloha network with a M/M/1/K buffer, as well as through simulations. We show that this effect can be understood in terms of an increased

number of packets reaching the MAC service, which leads to increased contention and consequently to a throughput drop. This result has implications for the provisioning of buffers for traffic with a delay/packet loss trade-off, such as interactive applications. We show that by appropriately configuring the channel access parameters of the protocol, latency can be controlled by tuning the buffer size according to the application requirements.

References

1. Asmussen, S.: Applied Probability and Queues, 2nd edn. Springer (2003)
2. Bertsekas, D.P., Gallager, R.G.: Data networks, vol. 2. Prentice-Hall (1987)
3. Bianchi, G.: Performance analysis of IEEE 802.11 distributed coordination function. IEEE Journal on Selected Areas in Communications 18(3), 535–547 (2000)
4. Carrig, B., Denieffe, D., Murphy, J.: Supporting first person shooter games in wireless local area networks. In: Proc. IEEE PIMRC, pp. 1–6 (September 2007)
5. Claypool, M., Claypool, K.: Latency and player actions in online games. Commun. ACM 49 (2006)
6. Duffy, K., Ganesh, A.J.: Modeling the impact of buffering on 802.11. IEEE Communications Letters 11(2) (February 2007)
7. Huang, K., Duffy, K.R., Malone, D.: On the validity of IEEE 802.11 MAC modeling hypotheses. IEEE/ACM Transactions on Networking 18(6), 1935–1948 (2010)
8. Little, J., Graves, S.: Little's law. In: Building Intuition. International Series in Operations Research & Management Science, vol. 115, pp. 81–100. Springer US (2008)
9. Liu, R.P., Sutton, G., Collings, I.B.: A new queueing model for QoS analysis of IEEE 802.11 DCF with finite buffer and load. IEEE Transactions on Wireless Communications 9(8), 2664–2675 (2010)
10. Malone, D., Duffy, K., Leith, D.: Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions. IEEE/ACM Transactions on Networking 15(1), 159–172 (2007)
11. Patras, P., Banchs, A., Serrano, P., Azcorra, A.: A Control-Theoretic Approach to Distributed Optimal Configuration of 802.11 WLANs. IEEE Transactions on Mobile Computing 10(6), 897–910 (2011)
12. Zhai, H., Kwon, Y., Fang, Y.: Performance analysis of IEEE 802.11 MAC protocols in wireless LANs. Wireless Communications and Mobile Computing 4(8), 917–931 (2004)