Modeling the Impact of Buffering on 802.11

Ken Duffy and Ayalvadi J. Ganesh

Abstract— A finite load, large buffer model for the WLAN medium access protocol IEEE 802.11 is developed that gives throughput and delay predictions. This enables us to investigate the impact of buffering on resource allocation. In the presence of heterogeneous loads, 802.11 does not allocate transmission opportunities equally. It is shown that increased buffering can help this inequity, but only at the expense of possibly significantly increased delays.

Index Terms— WLAN, IEEE 802.11, performance evaluation.

I. INTRODUCTION

BY design, in a network of equally loaded stations the IEEE 802.11 Distributed Coordination Function (DCF) Medium Access Controller (MAC) gives, in the long run, symmetric access. Each station gets an approximately equal number of access opportunities. However, in most deployments offered loads are asymmetric. For example consider the typical usage case: an infrastructure mode network where the Access Point (AP) has a wired back-haul. Load at the AP is approximately proportional to the number of stations in the network.

Using a finite load analytic model of 802.11 it is reported by Malone et al. [1] that in the presence of heterogeneous loads there is long-term inequity, with heavily loaded stations gaining more than their "fair share" of the bandwidth. For example, a 64kbps two-way voice conversation is considered in the presence of stations that always have a packet to send. With as few as 5 heavily loaded stations, the voice conversation's throughput is less than 32kbps. This inequity is due to the asymmetric nature of typical offered loads: data traffic, such as web and email, is typically bursty in nature while streaming traffic operates at relatively low rates and often in an on-off manner. To capture this analytically we require a finite load model. Note that short-term unfairness in 802.11 has previously reported (for example, see [2], [3], [4]), but it is fundamentally different to this long-term unfairness.

Stations with short buffers are treated in [1]. Here we extend the modeling paradigm developed in [5], [1] to give expressions for stations with large buffers and Poisson arrivals. This enables us to consider the impact of buffering on bandwidth-share inequity. With large buffers one expects that the inability to win access to the medium results in a backlog of packets awaiting transmission. This leads to an effectively higher offered load (in comparison to a short buffer where traffic that arrives while a packet awaiting access to the medium is lost) and thus a return towards a more equitable bandwidth

share. With a short buffer packet loss is a good quality-of-service indicator and delay is less important. With large buffers total delay (MAC plus queueing delay) is the most important performance indicator, so we provide an estimate of it. Note that these large buffer expressions for throughput and delay can be used in conjunction with the mesh network model proposed in [6].

II. PRELIMINARIES

As we extend the methodology in [5], [1] to treat an infinite buffer with Poisson arrivals, we start with a brief overview of the model. It is a mean-field Markov model of the sort introduced by Bianchi [7]. With a network of N stations we assume that each station $n \in \{1, \ldots, N\}$ has a fixed probability p_n of collision given it is attempting transmission, irrespective of its back-off stage. We describe the offered load of each station n by two probabilities, q_n and r_n, that are internal to the model. We will relate these to real-world offered load. When a station is in post-backoff or its count-down has completed and it is awaiting a packet, q_n denotes the probability a packet arrives to the MAC during an average slot time on the medium (which can be occupied by no station transmitting, a station successfully transmitting or a collision). The parameter r_n corresponds to the probability that immediately after a successful transmission a packet is available to the MAC. This is a generalization from [5], [1] where $q_n = r_n$, as when a station has no buffer these are the same, but in the presence of buffers they differ.

Under these assumptions, the back-off procedure forms an embedded (non real-time) Markov chain. Its stationary distribution can be calculated explicitly by the derivation described in [1] to give an expression for $\tau_n := \tau(p_n, q_n, r_n)$, the stationary probability that station n is attempting transmission in a slot. Temporarily dropping the n subscript,

$$\tau := \frac{1}{\eta} \frac{1}{1-r} \left(\frac{q^2 W_0}{(1-p)(1-(1-q)^{W_0})} - rq(1-p) \right), \quad (1)$$

where

$$
\begin{aligned}
\eta \quad := \quad & (1-q) + \frac{q^2 W_0 (W_0+1)}{2(1-(1-q)^{W_0})} \\
& + \frac{W_0+1}{2(1-r)} \left(\frac{q^2 r W_0}{1-(1-q)^{W_0}} + qp(1-r) - qr(1-p)^2 \right) \\
& + \frac{p}{2(1-r)(1-p)} \left(\frac{q^2 W_0}{1-(1-q)^{W_0}} - rq(1-p)^2 \right) \\
& \qquad\qquad \left(2W_0 \frac{1-p-p(2p)^{m-1}}{1-2p} + 1 \right),
\end{aligned}
$$

W_0 is the station's minimum contention window and $W_0 2^m$ is the station's maximum window size.

For given probabilities $\{(q_n, r_n)\}$, the conditional collision probabilities $\{p_n\}$ and transmission probabilities $\{\tau_n\}$ are completely determined by the solution of the fixed point equations that say that the probability a station doesn't experience

a collision given it is attempting transmission is the probability that no-one else is attempting transmission:

$$1 - p_n = \prod_{i \neq n} (1 - \tau_i(p_i, q_i, r_i)) \text{ for } n \in \{1, \ldots, N\}. \quad (2)$$

As the Markov chain doesn't evolve in real-time, to make real-time predictions we must determine the expected time between counter decrements (as given in [7]):

$$T := (1 - P_{tr})\sigma + P_{tr}P_s T_s + P_{tr}(1 - P_s)T_c, \quad (3)$$

where $P_{tr} = 1 - \prod_n (1 - \tau_n)$, $P_s = \sum_{i=1}^{n} \tau_n \prod_{j \neq i} (1 - \tau_j)/P_{tr}$, E is the time spent transmitting payload data (which for simplicity we assume is the same for packets from all stations; general expressions can be found in [1]), σ is the time for the counter to decrement, T_s is the time for a successful transmission and T_c is the time for a collision. For example, the throughput of station n is then $S_n = \tau_n (1 - p_n)E/T$.

III. RELATING OFFERED LOAD TO (q, r)

To make the model predictive we related the internal load parameters $\{(q_n, r_n)\}$ to real-world offered load. In [5], [1], a relation is given in the absence of buffers, so that $q_n = r_n$. With i.i.d. inter-arrivals with exponential distribution t_n of rate λ_n, the probability that no packet arrives during an average transition time in the Markov chain is $1 - q_n = \mathbf{P}(t_n > T) = \exp(-\lambda_n T)$, where T is given in equation (3). Thus for a given collection of arrival rates $\{\lambda_n\}$, one solves (2) for a range of $\{q_n\}$, identifying a collection such that $q_n = 1 - \exp(\lambda_n T)$ for all n.

Here we give a new relation based on an infinite buffer with Poisson arrivals. We relate the probability q_n to λ_n as above, but r_n no longer equals q_n. We treat each station as an M/G/1 queue, where the service time distribution G is the MAC delay to successful transmission. From the well known formula for the steady state probability that there is a packet in an M/G/1 queue after a packet transmission we determine r_n as a function of q_n and p_n. This reduces $\tau(p_n, q_n, r_n)$ to be a function of p_n and q_n. We must determine $\mathbf{E}(G)$. To do this we first consider the distribution $B(p)$ of the number of states in the Markov chain that pass for each packet prior to successful transmission given the conditional collision probability is p, which is approximately equal in distribution to

$$X_0 + Y_1 X_1 + Y_1 Y_2 X_2 + \ldots \quad (4)$$

where $\{X_n\}$ forms an independent sequence with X_n uniformly distributed on $[0, 2^{\min(n,m)} W_0 - 1]$ and $\{Y_n\}$ is an i.i.d. sequence of Bernoulli random variables with $\mathbf{P}(Y_1 = 1) = p = 1 - \mathbf{P}(Y_1 = 0)$. We say approximately as (4) is an upper bound that ignores post-back off and assumes every packet experiences at least one count-down. It is shown in [1] that this is a good approximation. From (4) it is possible to show that

$$\mathbf{E}(B(p)) = \frac{W_0}{2(1 - 2p)}(1 - p - p(2p)^m). \quad (5)$$

The steady state probability that an M/G/1 queue has packet after a transmission is $\min(1, \lambda \mathbf{E}(G))$, and $\mathbf{E}(G) = \mathbf{E}(B(p))T$. Hence $r_n = \min(1, -B(p_n)\log(1 - q_n))$ and $\tau(p_n, q_n, r_n)$ is only a function of p_n and q_n. Thus, again, for
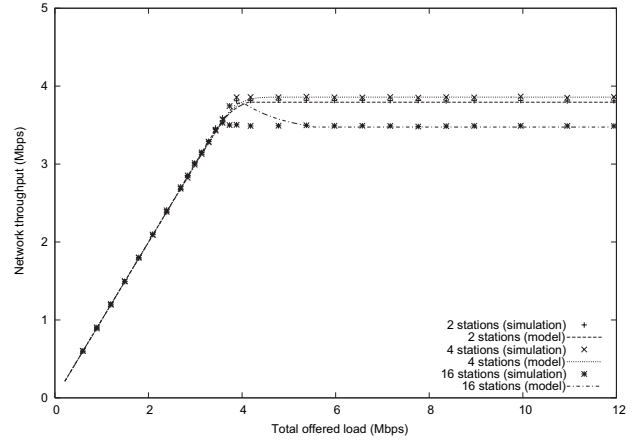

Fig. 1. Symmetric network throughput. Model predictions and NS simulation.

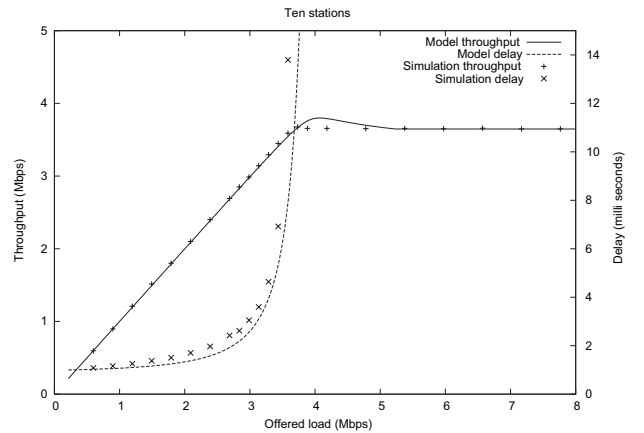Fig. 2. Symmetric network throughput and delay. Model predictions and NS simulations.

a given collection of arrival rates $\{\lambda_n\}$, one solves (2) for a range of $\{q_n\}$, identifying those for which $q_n = 1 - \exp(\lambda_n T)$ for all n. Let $\{p_n^\star\}$ denote this solution of (2). Once we know $\{p_n^\star\}$ it is possible to estimate the average queueing delay at station n by a standard formula (e.g. pg 237 of [8]). Using (4) a lengthy calculation gives

$$\mathbf{E}(B(p)^2) = \frac{p}{6(1-p)} + \frac{p^m}{(1-p)} \left[\frac{2^{2m}W_0^2}{3} - \frac{2^m W_0}{2} \right]$$
$$+ \frac{W_0^3}{3} \frac{(1-(4p)^m)}{1-4p} - \frac{W_0}{2} \frac{(1-(2p)^m)}{1-2p} + 2 \left[p^m \left(\frac{2^m W_0 - 1}{2(1-p)} \right)^2 \right.$$
$$- \frac{W_0(1-(2p)^m)}{2(1-2p)^2} + \frac{1-p^m}{2(1-p)^2} + \frac{W_0^2}{4(1-2p)} \frac{(1-(4p)^m)}{1-4p}$$
$$\left. - \frac{W_0^2(2p)^m}{4(1-2p)}(2^m - 1) - \frac{W_0}{4(1-p)} \frac{(1-(2p)^m)}{1-2p} + \frac{W_0^2(2p)^m}{4(1-p)}(2^m - 1) \right].$$

Then, if $\lambda_n \mathbf{E}(B(p_n^\star)) < 1$, the average delay is

$$\frac{\lambda_n \mathbf{E}(B(p_n^\star)^2)T^2}{2(1 - \lambda_n \mathbf{E}(B(p_n^\star))T)}.$$

If $\lambda_n \mathbf{E}(B(p^\star)) > 1$, the queue is unstable and the average delay is infinite. The mean MAC delay at station n is $\mathbf{E}(B(p_n^\star))T$.

IV. MODEL VALIDATION

Although it is not possible to present extensive validation due to space constraints, Figure 1 gives a good indication of the model's throughput accuracy. All stations have 560 byte
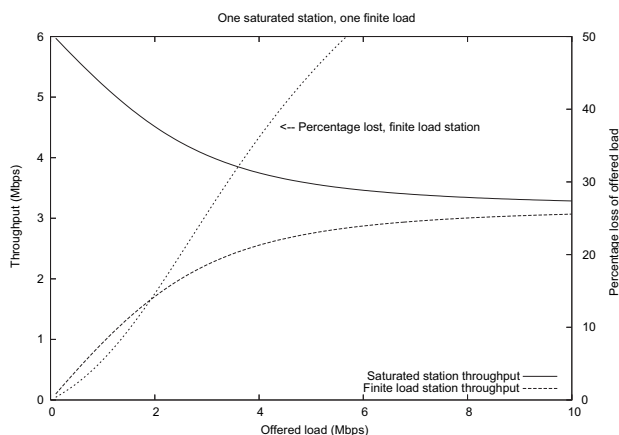
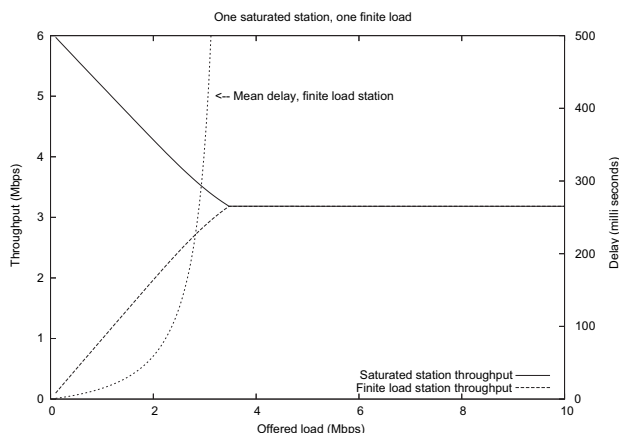Fig. 3. Asymmetric network throughput and loss. Small buffer predictions.

Fig. 4. Asymmetric network throughput and delay. Big buffer predictions.

Fig. 5. Throughput and delay for VoIP/TCP network. Model predictions.

packets and we use standard 11Mbps 802.11b parameterization, chosen so that direct comparison is possible with the short buffer results in [5]. The pre-saturation peak reported in [5] for short buffers, although present, is less pronounced and slightly overestimated by the model for large number of stations. For a network with 10 stations, Figure 2 plots throughput and delay predictions versus simulation results. Note the sudden, sharp climb in delay as a function of offered load, which takes place near peak throughput.

V. FAIRNESS

Figure 3 shows throughput and loss versus offered load in an asymmetric network with one saturated station and one with finite load, where the finite load station has a short buffer. The finite load station fails to get its fair share, except at high loads and experiences massive loss even at low loads. Figure 4 shows station throughput and mean queueing delay for the same scenario, but with the finite load station having a large buffer. It is clear that buffering is a significant factor in enabling the lower-load station to grab its share of the bandwidth. However, increased buffering leads to a dramatic ramp up in delay near the point at which loss in the short buffer model is becoming unacceptable. We have seen qualitatively similar results for larger numbers of saturated stations.

Extra buffer space is not a panacea due to the possibility of delay sensitive traffic. For example, consider a two-way voice conversation with each half on a distinct stations that transmits
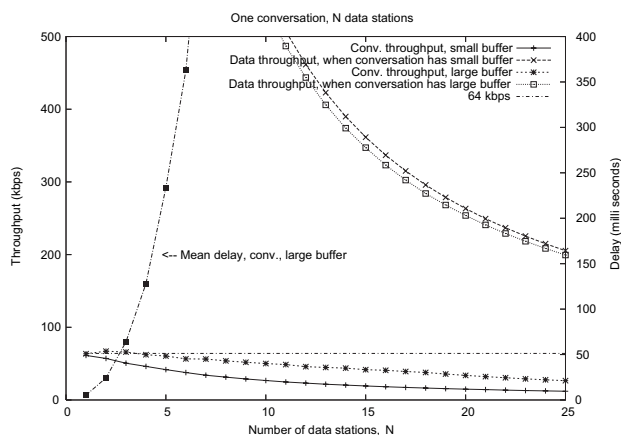
at 64k when active. We model the voice by a pair of stations with 100 byte packet Poisson traffic streams whose rate gives a 32k offered load. We model data stations as saturated, always having 1500 byte packets to send. Figure 5 shows model predictions of throughput for the voice call and a data station, as a function of the number of data stations, for both large and small buffers. Average delay for the voice is shown for the large buffer. Buffering increases the throughput of the voice conversation, but with as few as five data stations the delay is unmanageable for a real-time application. For delay sensitive traffic, prioritization using the faculties of 802.11e seems more appropriate.

VI. CONCLUSIONS

Increasing buffering to enable a station grab its "fair" share of the bandwidth in an asymmetrically loaded network is a double-edged sword. It aids with bandwidth share, but does so at the penalty of significantly increased delays. Thus increased buffering is probably not a suitable solution for time-constrained traffic such as VoIP and it is necessary to use the feature set of 802.11e.

REFERENCES

[1] D. Malone, K. Duffy, and D. J. Leith, "Modeling the 802.11 distributed coordination function in non-saturated heterogeneous conditions," to appear in *IEEE/ACM Trans. Network*, 2007.
[2] C. E. Koksal, H. Kassab, and H. Balakrishnan, "An analysis of short-term fairness in wireless media access protocols," in *Proc. ACM SIGMETRICS*, June 2000.
[3] G. Berger-Sabbatel, A. Duda, M Heusse, and F. Rousseau, "Short-term fairness of 802.11 networks with several hosts," in *Proc. MWCN 2004*.
[4] A. Kumar, "Analysis and optimisation of IEEE 802.11 wireless local area networks," in *Proc. WiOpt 2005*.
[5] K. Duffy, D. Malone, and D. J. Leith, "Modeling the 802.11 distributed coordination function in non-saturated conditions," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 715–717, Aug. 2005.
[6] K. Duffy, D. J. Leith, T. Li, and D. Malone, "Modeling 802.11 mesh networks," *IEEE Commun. Lett.*, vol. 10, no. 8, pp. 635–637, Aug. 2006.
[7] G. Bianchi, "Performance analysis of IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, March 2000.
[8] S. Asmussen, "Applied probability and queues," vol. 51 of *Applications of Mathematics*. New York: Springer-Verlag, second edition, 2003, Stochastic Modelling and Applied Probability.