New Insights from a Fixed Point Analysis of Single Cell IEEE 802.11 WLANs*

Anurag Kumar[1], Eitan Altman[2], Daniele Miorandi[2] and Munish Goyal[1]

***Abstract*— We study a fixed point formalisation of the well known analysis of Bianchi. We provide a significant simplification and generalisation of the analysis. In this more general framework, the fixed point solution and performance measures resulting from it are studied. Uniqueness of the fixed point is established. Simple and general throughput formulas are provided. It is shown that the throughput of any flow will be bounded by the one with the smallest transmission rate. The aggregate throughput is bounded by the reciprocal of the harmonic mean of the transmission rates. In an asymptotic regime with a large number of nodes, explicit formulas for the collision probability, the aggregate attempt rate and the aggregate throughput are provided. The results from the analysis are compared with *ns2* simulations, and also with an exact Markov model of the back-off process. It is shown how the saturated network analysis can be used to obtain TCP transfer throughputs in some cases.**

Keywords: wireless networks, CSMA/CA, performance of MAC protocols

I. Introduction

We are concerned in this paper with the situation in which there are several IEEE 802.11 compliant nodes within such a distance of each other that only one transmission can be sustained at any point of time. We call these *single-cell* networks. Our discussion covers *ad hoc networks*, and also *infrastructure networks*, in which an AP acts as a conduit between the wireless network and a wired "infrastructure." Our analysis is limited to the situation in which all nodes use the RTS/CTS based distributed coordination function (DCF) without the QoS extensions (as in IEEE 802.11e) (but see [8] for our extensions of the work in the present paper).

Each node may have several physical *connections or associations* with several other nodes. On each such connection the sustainable *physical* transmission rate may be different. Between each such pair of nodes there are *flows* whose throughput performance we are concerned with. It is assumed throughout this paper that all flows are infinitely back-logged at their transmitters; i.e., there are always packets to transmit when a node gets a chance to do so.

In such a scenario, we are interested in obtaining quantitative formulas and qualitative insights via a stochastic analysis of the way that the IEEE 802.11 CSMA/CA protocol allocates the wireless medium to the node transmitters. Our approach is to begin with a key approximation made by Bianchi [2]. This leads to a fixed point equation, which can be expected to characterise the operating points of the system. This fixed point equation is our point of departure. We simplify and generalise the analysis leading to the fixed point equation. We then establish a simple, and practically appealing, condition for the uniqueness of the fixed point in this more general framework. Some simple observations lead to throughput formulas for the overall network and for the individual flows. These formulas allow us to recover the well known observation that the slowest transmission rate dominates the throughput performance. We also analyse the fixed point in the asymptotic regime of a large number of nodes and find explicit formulas for the collision probability, the channel access rate and the network throughput. A key parameter in the protocol is the back-off multiplier, whose default value in the IEEE 802.11 MAC standard is 2; our asymptotic analysis provides some insights into the role of the back-off multiplier.

We provide *ns2* simulation results for the collision probabilities and compare these with results obtained from the fixed point analysis. We also provide results from an exact Markov chain model for the back-off process and also compare these results with those from the fixed point analysis.

As already pointed out, the above described modeling assumes that there are always packets backlogged on every connection. Such a *saturation assumption* is a common simplification and is useful in the following ways. In some situations it has been formally proved (see, for example, [4]) that the saturation throughput provides a sufficient condition for stability of the queues; i.e., if at each queue the arrival rate is less than the saturation throughput then the queues will have a proper, joint stationary distribution. In this paper we also apply the saturation throughput analysis to provide an analysis for TCP controlled file transfer throughputs in certain local area network scenarios.

The most popular model for IEEE 802.11 networks, and one that has led to many applications and extensions, is the one reported in [2]. Another analysis, that also incorporates the feature of adapting the back-off parameters, has been reported in [3]. The recent paper [1] is one of the many that have reported a throughput "anomaly" in IEEE 802.11 networks; i.e., if the network has low speed connections, even the high speed connections experience throughput no better than what is obtained by the low speed connections.

The paper is organised as follows. In Sec II we provide the key observation and approximation on which the analysis is based. In Sec III we analyse the back-off process in a fairly general setting. The fixed-point equation is provided in Sec IV and analysed in Sec V; a validation through an exact

*This work was supported by the Indo-French Centre for Promotion of Advanced Research (IFCPAR) under research contract No. 2900-IT.

[1]ECE Department, Indian Institute of Science, Bangalore, INDIA
[2]INRIA, Sophia-Antipolis, FRANCE

solution of a Markov model is given in Sec V-B. In Sec VI the throughput formulas are provided. The asymptotic analysis is developed in Sec VII. An application of the results to the analysis of TCP is given in Sec VIII and the paper ends with a concluding section. Some proofs are provided in-line and others are in the Appendix. Many details, not provided in this paper (including one proof), can be found in the report [5].

II. A Key Observation and an Approximation

A. *Sufficient to Analyse the Back-Off Process*

We begin by extracting from a description of the system the key modeling abstractions that will allow us to develop the analysis. Figure 1 shows the evolution of the system for 4 nodes; shown are the back-offs, the transmissions and collisions. In the IEEE 802.11 standard, the back-off durations are in multiples of a standardised time interval called a *slot* (e.g., 20 μs in IEEE 802.11b). However, this discrete nature of the back-offs does not affect the following argument. When a node completes its back-off (for example, node 1 is the first to complete its back-off in Figure 1), it seeks a reservation of the channel by sending an RTS packet. If no other node completes its back-off before hearing this transmission then the RTS effectively reserves the channel for the first node. There follows a CTS from the intended recipient of the RTS, and then there follows a packet transmission and a MAC level ACK. This ends the reservation period and the node that transmitted the packet samples a new back-off interval. Note that we assume throughout that nodes always have packets to transmit; i.e., *all the transmission queues are saturated*.

If the RTS collides with that of another node (note that we do not model the phenomenon of packet capture), then after fully transmitting their RTSs each node waits for a time interval DIFS before returning to the back-off state. For example, in Figure 1 nodes 2 and 4 collide after the first two attempts (by nodes 1 and 3, respectively) are successful. The other nodes, not involved in the collision, listen to the channel activity until the end of the RTS transmissions, and then also start their DIFS timers. Thus after a collision all nodes resume their back-off phases after an amount of time equal to the transmission time of an RTS plus a DIFS.

If attempts to send the packet at the head-of-the-line (HOL) meet with several successive failures, this packet is discarded. By our assumption of saturated queues, there is always another packet waiting to be sent by the upper layers: either the same packet or the next one in line.

We see from the figure that when any node has reserved the channel or whenever there is a collision, all other nodes freeze their back-off timers. We also notice that the evolution of the channel activity after an attempt is deterministic. It is either the time taken for a transmission or for a collision. If there is a transmission then the time depends on which node captures the channel. The latter dependence comes about because the transmission time of a packet depends on the transmission rate and hence on the transmitting node.

Since all nodes freeze their back-offs during channel activity, the total time spent in back-off up to any time t, is the *same* for every node. With this observation, let us now look at Figure 2 which shows the back-offs of Figure 1 with the channel activity removed. Thus in this picture "time" is just the cumulative back-off time at each node. In the IEEE 802.11 standard the back-offs are multiples of the slot time. A success occurs if a single back-off ends at a slot boundary, and a collision occurs when two or more back-offs end at a slot boundary. The nodes could have different back-off parameters (the mean back-off intervals, how these are varied in response to collisions and successes, and the number of retries of a packet). It is clear, however, that the (random) sequence in which the nodes seek turns to access the channel and whether or not each such attempt succeeds depends only on the back-off process shown in Figure 2. It is therefore sufficient to analyse the back-off process in order to understand the channel allocation process. The saturation assumption is crucial here since, with this assumption, we do not have to take care of any external packet arrivals that may occur during channel activity periods.

Thus, in summary, we can delete the channel activity periods, and we are left with a "conditional time" which we will call *back-off time*. We will analyse the back-off process conditioned on being in back-off time. It will then be shown how this analysis can be used to yield the desired performance measures over all time.

B. *A Key Approximation*

Throughout the rest of the paper *we assume that all the nodes use the same back-off parameters*. Hence the back-off process shown in Figure 2 is symmetric over the nodes. We call this the **homogeneous case** to distinguish it from the **nonhomogeneous** case in which different nodes may use different back-off parameters, as, for example, proposed in the IEEE 802.11e standard (see [7]).

In Figure 2 we also show the aggregate sequence of successes and collisions. In general, this is a complex process, and it is also clear that the success and collision processes of the various nodes are coupled and strongly correlated. In Sec V-B we will describe an exact Markov chain model for the joint backoff process of the nodes, but this model is analytically intractable. The following key approximation is made in [2].

The Decoupling Approximation: Let β denote the long run average back-off rate (*in back-off time*) for each node. By the fact that all nodes use the same back-off parameters, and by symmetry, it is assumed that all nodes achieve the same value of β. Let there be n contending transmitters, and consider a given node. The key approximation is to assume that the aggregate attempt process of the other $(n-1)$ nodes is independent of the back-off process of the given node. In IEEE 802.11 the back-off evolves over slots, hence a discrete time model (embedded at slot boundaries) can be adopted. Then the approximation says that if the attempt rate per node is β attempts per slot ($0 \leq \beta \leq 1$), then from the point of view of the given node the number of attempts by the other nodes in successive slots are i.i.d. Binomial random variables with parameters $(n-1)$ and β. It might be expected that such

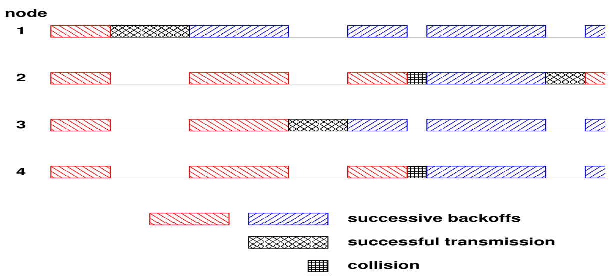

Fig. 1. The evolution of the back-off periods and channel activity for four nodes. It can be seen that back-offs are interrupted by channel activity, i.e., packet transmissions and RTS collisions.

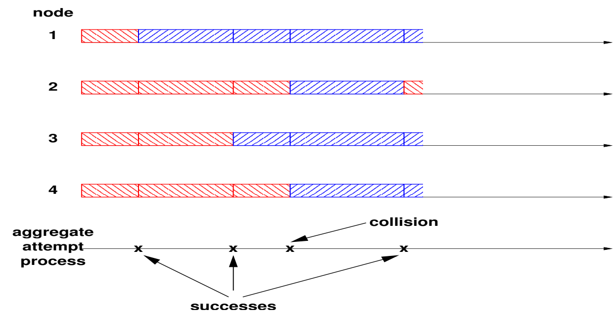

Fig. 2. After removing the channel activity from Figure 1 only the back-offs remain. These determine the scheduling of channel access. At the bottom is shown the aggregate attempt process on the channel, with three successes and one collision.

a decoupling approximation should work well when there is a large number of transmitters accessing the channel.

III. Analysis of the Back-Off Process

We generalise the back-off behaviour of the nodes, and define the following back-off parameters.

$K :=$ At the $(K+1)$th attempt either the packet succeeds or is discarded

$b_k :=$ The mean back-off duration (in slots) at the kth attempt for a packet, $0 \leq k \leq K$

Since we are limiting ourselves to the homogeneous case, these parameters are the same for all the nodes.

In Figure 3 we show the evolution of the back-off process for a single node. There are R_j attempts until success for the jth packet (no case of a discarded packet is shown in this diagram), and the sequence of back-offs for the jth packet is $B_j^{(i)}, 0 \leq i \leq R_j - 1$. Thus the total back-off for the jth packet is given by $X_j = \sum_{i=0}^{R_j-1} B_j^{(i)}$ with $\mathsf{E}\left(B_j^{(i)}\right) = b_i$. We observe that the sequence $X_j, j \geq 1$, are renewal life times. Hence, viewing the number of attempts R_j for the jth packet as a "reward" associated with the renewal cycle of length X_j, we obtain from the renewal reward theorem that the back-off rate is given by $\mathsf{E}(R)/\mathsf{E}(X)$. Now let γ be the collision probability seen by a node, i.e.,

$$\gamma := \mathsf{Pr}\,(\text{an attempt by a node fails because of a collision})$$

Since the back-off behaviour of all the nodes is the same, the collision probability is the same for all the nodes. By the approximation made in Sec II, the successive collision events are independent. It is then easily seen that

$$\begin{aligned} \mathsf{E}(R) &= 1 + \gamma + \gamma^2 \cdots + \gamma^K \\ \mathsf{E}(X) &= b_0 + \gamma b_1 + \gamma^2 b_2 + \cdots + \gamma^k b_k + \cdots + \gamma^K b_K \end{aligned}$$

which yields the following formula for the attempt rate for a given collision probability γ

$$G(\gamma) := \frac{1 + \gamma + \gamma^2 \cdots + \gamma^K}{b_0 + \gamma b_1 + \gamma^2 b_2 + \cdots + \gamma^k b_k + \cdots + \gamma^K b_K} \quad (1)$$

Note that, since the back-off times are in slots, the attempt rate $G(\gamma)$ is in *attempts per slot*.

Remarks 3.1:

1) Note that the *distribution* of the back-off durations does not matter. Also, observe that the above analysis remains unchanged whether the back-off distributions are discrete (i.e., the back-offs evolve over slots) or are continuous.

2) It is easily seen that the back-off model considered in [2] yields

$$G(\gamma) = \frac{2(1-2\gamma)}{(1-2\gamma)(CW_{\min} \pm 1) + \gamma CW_{\min}(1-(2\gamma)^m)}$$

attempts per slot, which is the same as in the paper [2]. Note that the $\pm$ alternatives arise depending on whether we take the back-off to be uniformly distributed over $[1, 2, \cdots, CW]$ or over $[0, 1, \cdots, CW-1]$. Evidently, the uniform distribution of back-off durations plays no role in the final results in [2].

3) A more detailed evolution of the back-off process in Figure 3 is shown in Figure 4, where at each time t the residual back-off duration $Y(t)$ is also shown. The process $Z(t)$ is the back-off *stage* the node is in. Thus if $K = 7$, $Z(t) = 3$, and $Y(t) = 5$, then after 5 time units the current back-off ends. If there is a collision, $Z(t)$ changes to 4 and a back-off with mean b_4 is sampled from the specified back-off distribution (uniform in the standard). If $Z(t) = 7$ then at the end of the current back-off, irrespective of whether there is a collision or a success, the next back-off has mean b_0, and is sampled from the specified distribution. It is clear that the process $(Z(t), Y(t))$ is Markov. The point is that it is not necessary to analyse this Markov chain, which is essentially what is done in [2]. Let $Z_k, k \geq 0$, denote the process $Z(t)$ embedded at the attempt instants (the instants corresponding to the vertical sides of the triangles in Figure 4). Then Z_k is an embedded Markov chain. Further, Z_k and the successive back-off intervals (the bases of the triangles) constitute a Markov renewal process. It is well known that for Markov renewal processes event rates and time probabilities are insensitive to distributions of life-times. It should thus be clear why one can directly obtain the formulas above without needing to go

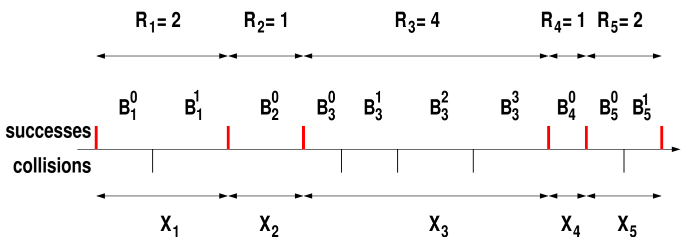

Fig. 3. Evolution of the back-offs of a node. Each attempted packet starts a new back-off "cycle."

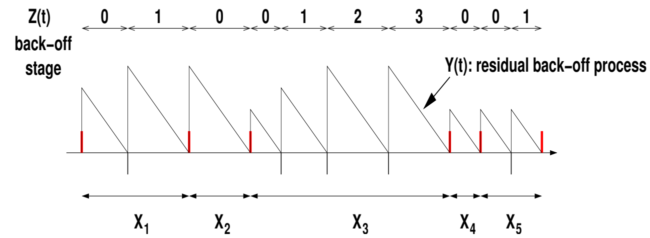

Fig. 4. Evolution of the back-off stage ($Z(t)$) and the residual back-off time ($Y(t)$) for the case in which the back-offs are continuous variables.

through the analysis of the Markov chain in [2], and also why the results are insensitive to the back-off distribution.

IV. The Fixed Point Equation

Focusing on the back-off and attempt process of a node, and being given the collision probability γ the attempt rate is provided by $G(\gamma)$ in Eq 1. It is important to recall that in the present discussion all rates are conditioned on being in the back-off periods. Later we will see how to incorporate the channel activity periods. Now if all nodes have the same back-off parameters, they will all see the same average collision probability, γ, and hence will have the same attempt rate. If the attempt rate (or probability) of each node per slot is $\beta, 0 \leq \beta \leq 1$, then, conditioning on an attempt of the given node, the probability of this attempt experiencing a collision is the probability that any of the other nodes attempts in the same slot. Under the decoupling approximation, the number of attempts made by the other nodes is binomially distributed with parameters β and $n-1$. Under the approximation, the number of attempts in successive form an i.i.d. sequence. The probability of collision of an attempt by a node is given by

$$\Gamma(\beta) \quad := \quad 1-(1-\beta)^{(n-1)} \tag{2}$$

We will show later in the paper that under a certain asymptotic regime the aggregate attempt rate $n\beta$ converges to a positive value as $n \to \infty$. Then (motivated by the binomial to Poisson convergence theorem) for a large number of nodes, it is reasonable to model the attempt process of the other nodes (with respect to a given node) as a sequence of i.i.d. batches (at slot boundaries) with the batch distribution being Poisson with mean $(n-1)\beta$. The collision probability under this model is then clearly given by

$$\Gamma(\beta) \quad := \quad 1-e^{-(n-1)\beta} \tag{3}$$

It is now natural to expect that the equilibrium behaviour of the system will be characterised by the solutions of the following fixed point equation

$$\gamma \quad = \quad \Gamma(G(\gamma)) \tag{4}$$

If this equation can be solved it will yield the collision probability, from which the attempt rate can be obtained using Eq 1. We will see in Sec VI that throughputs can be obtained once these quantities are determined.

V. Analysis of the Fixed Point Problem

Since $\Gamma(G(\gamma_k))$ is a composition of continuous functions it is continuous. We thus have a continuous mapping from $[0,1]$ to $[0,1]$. Hence by Brouwer's fixed point theorem there exists a fixed point in $[0,1]$. We next turn to uniqueness.

Lemma 5.1: $G(\gamma)$ is non-increasing in γ if $b_k, k \geq 0$, is a non-decreasing sequence.

Proof: Provided in the Appendix. ■

Theorem 5.1: $\Gamma(G(\gamma)) : [0,1] \to [0,1]$, has a unique fixed point if $b_k, k \geq 0$, is a nondecreasing sequence.

Proof: Since $\Gamma(\beta)$ is non-decreasing in β and, by Lemma 5.1, $G(\gamma)$ is non-increasing in γ, it follows that $\Gamma(G(\gamma))$ is non-increasing in γ. The fixed point must therefore be unique, since multiple fixed points will lead to a contradiction to the non-increasing property of $\Gamma(G(\gamma))$. ■

Remarks 5.1:

(1) We observe that in the IEEE 802.11 standard the sequence b_k is non-decreasing. Hence for the practical system there will be a unique fixed point.

(2) In the above discussion we have only considered *balanced* fixed points, i.e., ones in which all the nodes have the same value of collision probability γ. It is possible, however, under the decoupling approximation, to set up a system of fixed point equations for *unbalanced* fixed points, i.e., ones in which the collision probability of node j is γ_j, with these values being possibly different for different j. By symmetry we expect that the long run average operating point of the system will correspond to a balanced fixed point. However, in [8] we have shown that in general there can also exist unbalanced fixed points, which suggest *multistability*, and indeed simulations reveal that in such cases there is serious short term unfairness. In [8] we also provide a sufficient condition for there to be no unbalanced fixed points. It turns out that the default IEEE 802.11 parameters satisfy these conditions. Thus in practice there will be a unique balanced fixed point and no unbalanced ones.

A. *Examples and Comparison with* ns2 *Simulations*

In Figure 5, we show plots of $\Gamma(G(\gamma))$ vs. γ for several parameters. Here $p=2$, as in the IEEE 802.11 standard. In the plot on the top we use the value $K=7$. In both the plots the initial mean back-off b_0 is 16 slots. The intersection of these plots with the "y=x" line corresponds to the fixed point. We

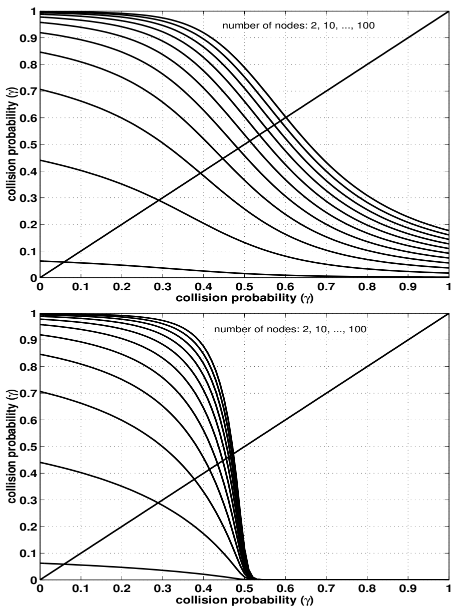

Fig. 5. Plots of $\Gamma(G(\gamma))$ vs. γ for two values of K (7 (top) and 100 (bottom)), $b_0 = 16$ slots, and multiplicatively increasing b_k with multiplier $p = 2$. For each K, plots are shown for number of nodes $n = 2, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100$.

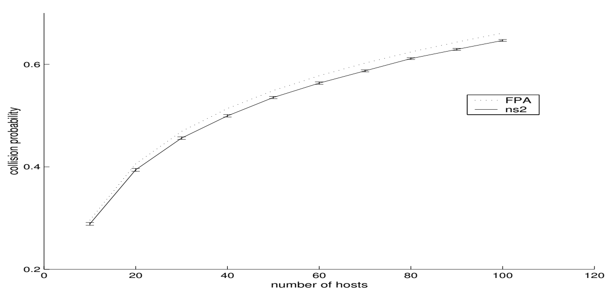

Fig. 6. Plot of collision probability versus number of nodes. Comparison of collision probability (γ) obtained from an *ns2* simulation (plot labeled *ns2*), and the fixed point analysis (plot labeled *FP*). 95% confidence intervals are shown for the values obtained from the *ns2* simulation. In the *ns2* simulation the default IEEE 802.11 parameters are used: the data rate is 11 Mbps and the control packet rate is 2 Mbps.

see that the collision probability increases with an increasing number of nodes. For $n \geq 30$, with $K = 7$, the collision probability is larger than with $K = 100$. This is because with larger K nodes are able to expand their back-off durations more and hence attempt less often. The collision probability for $n \leq 20$ is not sensitive to K for $K \geq 7$, since with $n \leq 20$ there are rarely more than 7 consecutive collisions.

It was reported in [2] that the fixed point analysis works well for IEEE 802.11 parameters. In Figure 6 we demonstrate this by plotting the collision probability obtained from the fixed point method and from an *ns2* simulation.

In all the *ns2* simulations presented in this paper we have used *ns2* version 2.26. The bugs present in the IEEE 802.11 code were patched by using an updated version of the code taken from the *ns2* snapshot dated January 5, 2004. Static routing was implemented by using *NOAH* code (dated November 2003), downloaded from the web site of J. Widmer, EPFL, (http://icapeople.epfl.ch/widmer/uwb/ns-2/noah/index.html). As can be seen, the fixed point analysis provides a good approximation for a wide range of values of the number nodes.

B. *Comparison with the Coupled Back-Off DTMC*

It can be seen that when the back-off durations are geometrically distributed, then the coupled evolution of the back-offs of the nodes, as shown in Figure 2, is exactly modeled by a discrete time Markov chain (DTMC). Hence if the decoupling approximation works well, it should be able to match the results obtained from this DTMC. We now turn to this question. We proceed with the following assumptions: (i)The number of nodes $n \geq 2$, (ii) Exponential back-off with multiplier $p > 1$, i.e., $b_k = p^k b_0, 1 \leq k \leq K$, (iii) Back-off durations are geometrically distributed, or, equivalently (with the b_k expressed in number of slots), when a node is in back-off stage k, it attempts in the next slot with probability $\frac{1}{b_k}$. We only need to consider the system back-off periods, and we index the slots in back-off time by $t = 0, 1, 2, \cdots$.

It is convenient to work with the process that counts the number of nodes in each back-off stage. This will be a $K+1$ dimensional process for any number of nodes. Define the number of nodes in the back-off stage $k \in \{0, 1, \cdots, K\}$ in slot t to be $M_k^{(n)}(t)$. Let $\mathbf{M}^{(n)}(t)$ denote the vector random process with components $M_k^{(n)}(t)$. From the foregoing, it is clear that $\mathbf{M}^{(n)}(t)$ is a Markov process taking values in the set $\mathcal{M}^{(n)} := \{\mathbf{m} : m_k \text{ nonnegative integers}; \sum_{k=0}^{K} m_k = n\}$.

Theorem 5.2: [5] For $b_0 > 1$, and $p > 1$, the DTMC $\mathbf{M}^{(n)}(t)$ on $\mathcal{M}^{(n)}$ is irreducible. ■

It follows that under the conditions $b_0 > 1$ and $p > 1$, the DTMC $\mathbf{M}^{(n)}(t)$ is positive recurrent. Let $\boldsymbol{\pi}^{(n)}$ denote the stationary probability measure on $\mathcal{M}^{(n)}$.

For small values of K (e.g., 1 or 2) $\boldsymbol{\pi}^{(n)}$ can be numerically computed. Now given $\boldsymbol{\pi}^{(n)}$, the collision probability γ can be obtained in a straight forward manner (see [5]). Sample results are shown in Table I. Results are shown for $K = 1$ and $K = 2$, and $b_0 = 16$. It can be seen that the fixed point analysis approximates the collision probability very well.

VI. Calculating Throughputs

We make two key observations. The first is demonstrated by Figure 7. Because of the i.i.d. batch binomial assumption

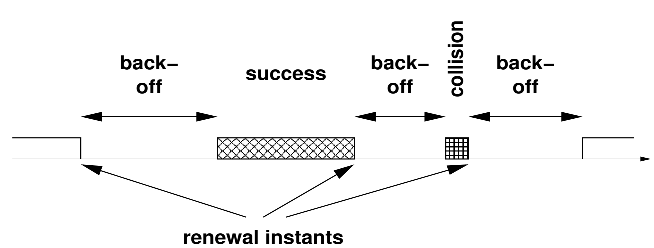

Fig. 7. The aggregate process of back-offs and channel activity

No. of Nodes	DTMC ($K=1$)	FPA ($K=1$)	DTMC ($K=2$)	FPA ($K=2$)
2	0.0598	0.0592	0.0595	0.0587
3	0.1111	0.1105	0.1088	0.1078
4	0.1568	0.1563	0.1510	0.1500
5	0.1983	0.1979	0.1879	0.1870
6	0.2365	0.2362	0.2209	0.2202
7	0.2720	0.2718	0.2508	0.2502
8	0.3052	0.3050	0.2782	0.2778
9	0.3363	0.3362	0.3036	0.3033
10	0.3657	0.3656	0.3272	0.3270
11	0.3933	0.3933	0.3494	0.3493
12	0.4196	0.4195	0.3703	0.3702
13	0.4444	0.4444	0.3900	0.3900
14	0.4680	0.4680	0.4088	0.4088
15	0.4905	0.4905	0.4266	0.4266
16	0.5119	0.5119	0.4436	0.4436
17	0.5323	0.5323	0.4598	0.4599
18	0.5518	0.5518	0.4754	0.4755
19	0.5703	0.5703	0.4903	0.4904
20	0.5881	0.5881	0.5046	0.5048

TABLE I

COLLISION PROBABILITIES: DTMC AND FIXED POINT ANALYSIS (FPA); $K=1$ AND $K=2$, AND $b_0=16$.

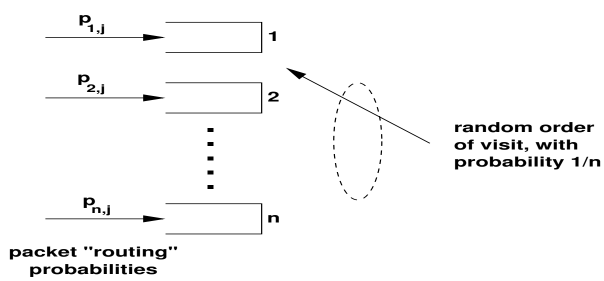

Fig. 8. The n transmitters are served in random order with equal probability for each node.

on the aggregate attempt process, the instants at which a successful transmission or a collision ends are renewal instants. Each such instant is followed by a time until the next attempt, followed by a collision or a success, and so on. The second observation is that since all the nodes follow the same back-off process, each node has an equal probability of winning the allocation "race." With this in mind we can now discard the back-off times and focus only on the times when an attempt is made and on the intervening channel activity. A successful attempt leads to the channel being allocated to one of the n contending nodes with equal probability. Hence in a saturated system, in order to compute the amount of time the channel will be allocated to a node, we only need to know the identity of the packet that will be found at the head-of-the-line if the channel is allocated to the node.

Consider the model shown in Figure 8. The nodes are visited in random order with equal probability. Each node receives an *open loop* stream of packets. There are m_i streams being handled by node i. These are indexed by $1 \leq j \leq m_i$; these would represent m_i flows from node i to some of the other nodes. We can thus use the term "flow (i,j)".

By "open-loop" we mean that packets arrive to the node and have to be delivered; there are no acknowledgement and flow control as in TCP controlled traffic. A fraction $p_{i,j}$ of the packets at node i belong to stream $j, 1 \leq j \leq m_i$. Since the node is saturated there is always a packet at the head-of-the-line when the channel is allocated to any node, and $p_{i,j}$ is the probability that the packet is from flow j. Let us define the packet length of flow (i,j) to be $L_{i,j}$ and the physical transmission rate for flow (i,j) to be $C_{i,j}$ *bits per slot*.

In addition, we define,

$T_o :=$ is the fixed overhead with a packet transmission in slots (e.g., IEEE 802.11b: $T_o = 52$ slots)

$T_c :=$ is the fixed overhead for an RTS collision in slots (e.g., IEEE 802.11b: $T_c = 17$ slots)

The above two observations, the traffic model described above, and the parameters listed above, lead immediately to the expression in Eq 5 for the saturation throughput of flow (i,j) (in bits per slot) given the collision probability γ and the per node attempt rate β.

The formula follows from the renewal reward theorem. The mean renewal time (see Figure 7) is the mean time until an attempt, plus the mean time for channel activity; i.e., a transmission or a collision. The term $\frac{1}{1-(1-\beta)^n}$ is the mean time until an attempt and assumes that the aggregate attempt process is binomial. When there is an attempt the channel is allocated to node i (with probability $\frac{\beta(1-\beta)^{n-1}}{1-(1-\beta)^n}$), else there is a collision, for which the channel will be busy for the time T_c. If the channel reservation succeeds, then the head-of-the-line packet at node i is of flow (i,j) with probability $p_{i,j}$, and transmitting this takes the time $\frac{L_{i,j}}{C_{i,j}} + T_o$. The mean reward during the cycle is $\frac{\beta(1-\beta)^{n-1}}{1-(1-\beta)^n} p_{i,j} L_{i,j}$, thus yielding the displayed expression. Canceling the term $1-(1-\beta)^n$, the formula simplifies to the expression in Eq 6.

A. Low Speed Transmitters Bound All Throughputs

It has been observed (see, for example, [1]) that when there are several flows with different physical transmission rates then the throughput of all the flows is bounded by the slowest transmission rate. We can examine this observation using Eq 6.

If 2 nodes i_1 and i_2 are such that for some $j_1, 1 \leq j_1 \leq m_1$, and $j_2, 1 \leq j_2 \leq m_2$, $p_{i_1,j_1} L_{i_1,j_1} = p_{i_2,j_2} L_{i_2,j_2}$ then it follows from Eq 6 that $\theta_{i_1,j_1}(\gamma,\beta) \leq \min\{C_{i_1,j_1}, C_{i_2,j_2}\}$ and $\theta_{i_2,j_2}(\gamma,\beta) \leq \min\{C_{i_1,j_1}, C_{i_2,j_2}\}$, i.e., the flow with the lower physical rate will bound the throughput of both.

Remark: The above analysis points to an important observation. Suppose we are interested in achieving flow throughputs that are proportional to their physical link rates; i.e., $\theta_{i,j} = \nu C_{i,j}$ for some ν. It has been suggested in previous literature that this can be achieved by appropriately choosing the packet lengths. We notice from Eq 6 that the desired throughput proportionality can be achieved only be making $L_{i,j}$ proportional to $\frac{C_{i,j}}{p_{i,j}}$, which requires knowledge of the $p_{i,j}$s, which may not be practicable.

Let us now consider a simpler situation with n nodes each being the transmitter for a single flow and all packet lengths being equal to L. Then the *total* network throughput is given

$$\theta_{i,j}(\beta) = \frac{\frac{\beta(1-\beta)^{n-1}}{1-(1-\beta)^n} p_{i,j} L_{i,j}}{\frac{1}{1-(1-\beta)^n} + \sum_{i=1}^n \left(\frac{\beta(1-\beta)^{n-1}}{1-(1-\beta)^n}\left(\left(\sum_{k=1}^{m_i} p_{i,k}\frac{L_{i,k}}{C_{i,k}}\right)+T_o\right)\right) + \left(\frac{(1-(1-\beta)^n-n\beta(1-\beta)^{n-1})}{1-(1-\beta)^n}\ T_c\right)} \quad (5)$$

$$\theta_{i,j}(\beta) = \frac{\beta(1-\beta)^{n-1} p_{i,j} L_{i,j}}{1 + \sum_{i=1}^n \left(\beta(1-\beta)^{n-1}\left(\left(\sum_{k=1}^{m_i} p_{i,k}\frac{L_{i,k}}{C_{i,k}}\right)+T_o\right)\right) + \left((1-(1-\beta)^n-n\beta(1-\beta)^{n-1})\ T_c\right)} \quad (6)$$

by

$$\Theta(\beta) = \frac{n\beta(1-\beta)^{n-1}L}{1 + \sum_{i=1}^{n}\left(\beta(1-\beta)^{n-1}\left(\frac{L}{C_i}+T_o\right)\right) + \left((1-(1-\beta)^n - n\beta(1-\beta)^{n-1})\ T_c\right)} \quad (7)$$

Since the denominator is bounded below by $\sum_{i=1}^n \left(\beta(1-\beta)^{n-1}\left(\frac{L}{C_i}+T_o\right)\right)$, it can be seen that

$$\Theta(\beta) \le \frac{1}{\frac{1}{n}\sum_{i=1}^n \frac{1}{C_i}} \le n \times \min_{1\le i\le n} C_i$$

i.e., *the total network throughput is bounded above by the reciprocal of the harmonic means of the physical bit rates of the n flows*. Thus, for example, if there are two flows with physical rates 2 Mbps and 4 Mbps then the total network throughput will be bounded by $\frac{2}{\frac{1}{2}+\frac{1}{4}} = 2.667$ Mbps. Also, with equal packet lengths, we see that this total throughput is shared equally among all the flows.

VII. An Asymptotic Analysis

If we numerically examine the fixed points (see [5]), we notice that the fixed points appear to be converging as K becomes large, and there is not much variation in them for $K \ge 15$. Thus we are motivated to analyse the fixed point for $K \to \infty$. A similar asymptotic analysis has also been carried out independently by Kwak et al in [6]; while their final results are the same as our Theorem 7.2, we have displayed an analytical form for the fixed point solution (see Theorem 7.1), and we derive our asymptotic results by taking a limit in this solution. Further, we also provide a relaxed fixed point iteration for computing the fixed point (see Sec VII-A).

To permit closed form analysis, let us take $b_0 = b$ slots, and $b_k = p^k \times b_0$, where $p \ge 1$; hence, by Theorem 5.1, a unique fixed point still exists. The multiplicative increase is in any case a part of the IEEE 802.11 standard; we are generalising to an arbitrary multiplier in order to study the impact of the value of this multiplier.

Assuming $\gamma < 1/p$, and taking $K \to \infty$, we see that

$$G(\gamma) = \frac{1}{b_o} \times \frac{1-p\gamma}{1-\gamma}$$

Note that the assumption that $\gamma < 1/p$ does not affect the fixed point analysis presented earlier, since we will see in Theorem 7.2 that the fixed point in the limit $K \to \infty$ is less than $1/p$.

Given γ, $G(\gamma)$ is the probability of attempt of any node. Then using the batch Poisson version of the collision probability in Eq 3, the fixed point equation becomes

$$\gamma = f(\gamma) \quad \text{where } f(\gamma) := 1 - \exp\left(-\frac{n-1}{b_0} \times \frac{1-p\gamma}{1-\gamma}\right) \quad (8)$$

In order to obtain compact expressions, let us define $\eta = \frac{n-1}{b_0}$.

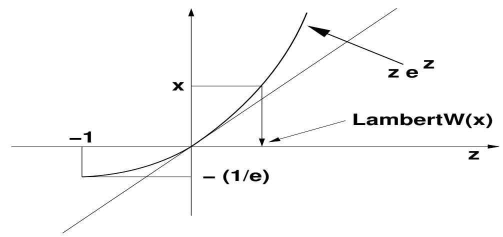

Fig. 9. The LambertW function is the inverse function of ze^z; notice that for $x \ge -\frac{1}{e}$, $LambertW(x) \le x$, with equality only for $x = 0$.

Theorem 7.1: The fixed point is of the form

$$\gamma(\eta) = \frac{LambertW(\eta(p-1)e^{\eta p}) - \eta(p-1)}{LambertW(\eta(p-1)e^{\eta p})}.$$

Remark: For $x \ge -1/e$, $LambertW(x)$ is defined as the inverse of the function ze^z; see Figure 9.

Proof: We proceed from Eq 8. Writing $\nu = 1-\gamma$, and using the definition of η, this equation can be rewritten as $\nu = \exp(-\eta p)\exp\left(\frac{\eta(p-1)}{\nu}\right)$. Multiplying both sides by $\eta(p-1)$, we obtain $\eta(p-1)\exp(\eta p) = \frac{\eta(p-1)}{\nu}\exp\left(\frac{\eta(p-1)}{\nu}\right)$. It follows from the definition of $LambertW(\cdot)$ that $\frac{\eta(p-1)}{\nu} = LambertW\Big(\eta(p-1)\exp(\eta p)\Big)$ from which the result follows by substituting $1-\gamma$ for ν. ■

A. A Relaxed Fixed Point Iteration

The fixed point $\gamma(\eta)$ can only be computed numerically. In this section we provide a relaxed fixed point iteration. With reference to Eq 8, and, with $\gamma_0 := 1/p$, consider the sequence of values generated by the iterations

$$\gamma_{k+1} = (1-\alpha)f(\gamma_k) + \alpha\gamma_k \quad (9)$$

where $0 < \alpha < 1$. Notice that $\alpha = 0$ corresponds to the usual fixed point iteration, which will converge if $f(\gamma)$ is a contraction. The above iteration is called a *relaxed* fixed point iteration. We will now provide a condition on α that will ensure that the iterates converge to the fixed point.

First of all, since $f(\gamma)$ is continuous, it is clear from the iteration in Eq 9 that if the sequence of iterates converge then they must converge to the fixed point. It is also clear that if, for each k, $\gamma_k \geq f(\gamma_k)$ then the sequence $\{\gamma_k\}$ is nonincreasing. This follows because $\gamma_{k+1} = (1-\alpha)f(\gamma_k) + \alpha\gamma_k \leq \gamma_k$ if and only if $f(\gamma_k) \leq \gamma_k$. Thus, since $\gamma_k \geq 0$, for the convergence of the sequence $\{\gamma_k\}$ it suffices to ensure that $\gamma_k \geq f(\gamma_k)$ for all k.

Now, it can easily be shown that the derivative of $f(\gamma)$ at $\gamma = 1/p$ is given by $D := -\frac{n-1}{b_0}\frac{p^2}{p-1}$, and that, for $\gamma_{k+1} \leq \gamma_k$

$$f(\gamma_{k+1}) - f(\gamma_k) \leq |D|\ (\gamma_k - \gamma_{k+1})$$

From this inequality we can see that to ensure $f(\gamma_k) \leq \gamma_k$, for all k, it is sufficient to ensure that, for all k, $|D|\ (\gamma_k - \gamma_{k+1}) \leq \gamma_{k+1} - f(\gamma_k)$. Using the iteration in Eq 9 this is equivalent to ensuring that, for all k, $|D|(1-\alpha)(\gamma_k - f(\gamma_k)) \leq \alpha(\gamma_k - f(\gamma_k))$. Hence it suffices that $|D|(1-\alpha) \leq \alpha$, or that $\alpha \geq |D|/(|D|+1)$. Thus, for example, with $n = 10$ nodes, $b_0 = 16$ slots, and $p = 2$, it relaxed fixed point iteration with α such that $\frac{2.25}{3.25} < \alpha < 1$ will yield the unique fixed point $\gamma(\eta)$.

B. *Taking n to ∞*

We now wish to take n to ∞ and study the limit of the fixed point solution obtained in Theorem 7.1. For this we need the following properties of the LambertW function.

Lemma 7.1:

1) For $a > 0$,

$$\lim_{x\to\infty} \frac{LambertW(axe^x)}{x} = 1 \qquad (10)$$

2) For $0 < a \leq 1$, $LambertW(axe^x) \leq x$
3) For $0 < a < 1$, the convergence in Eq 10 is from below.

Proof: Provided in the Appendix. ■

The following result is now obtained by applying Lemma 7.1 to the expression for γ in Theorem 7.1.

Theorem 7.2:

1) $\gamma(\eta) < 1/p$,
2) $\lim_{n\to\infty} \gamma(\eta) \uparrow 1/p$,
3) $\lim_{n\to\infty} n\beta \uparrow \ln(\frac{p}{p-1})$. ■

Remarks 7.1:

1) Theorem 7.2 provides explicit expressions for the collision probability and the fixed point for large K and a large number of nodes. We see that for large n the collision probability is directly related to the back-off multiplier p, and is the reciprocal of this multiplier.

2) We also see that $n\beta$, the mean attempt rate per slot, goes to $\ln(\frac{p}{p-1})$, and hence the attempt probability per node (during back-off periods) behaves like $O(\frac{1}{n})$. This lends some support to the original assumption that from the point of view of a node the attempt process of the other nodes can be viewed as an independent process with i.i.d. batch Poisson arrivals in successive slots.

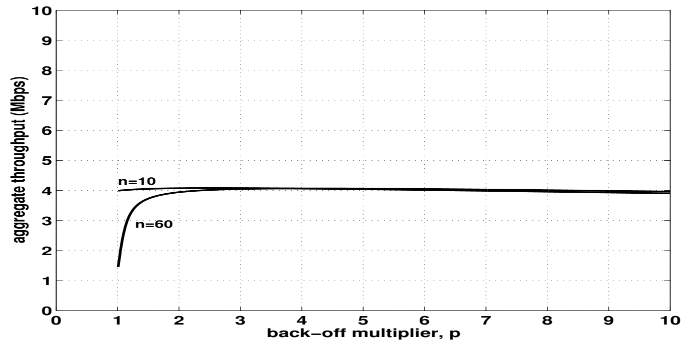

Fig. 10. Aggregate throughput plotted vs. the back-off multiplier p for two values values of n. The network parameters are $K = 10$, $b_0 = 16$ slots, data packet length 8000 bits, packet overhead 592 bits, slot time 20 μs, transmission rate for all flows 11 Mbps, fixed (rate independent) data packet transmission overhead 52 slots, collision overhead 17 slots.

C. *Asymptotic Aggregate Throughput*

Let us now consider n nodes handling n flows with all the flows having the same transmission rate, C. The aggregate throughput of the network is given by (compare with Eq 7)

$$\Theta(\beta) = \frac{n\beta e^{-n\beta}L}{1 + \left(n\beta e^{-n\beta}\left(\frac{L}{C} + T_o\right)\right) + \left(\left(1 - e^{-n\beta} - n\beta e^{-n\beta}\right)T_c\right)}$$

We infer from this equation that, as $n \to \infty$ the aggregate throughput converges to

$$\tau(p) := \frac{(1-\frac{1}{p})L}{\frac{1}{\ln(\frac{p}{p-1})} + (1-\frac{1}{p})(\frac{L}{C} + T_o) + \left(\frac{\frac{1}{p}}{\ln(\frac{p}{p-1})} - (1-\frac{1}{p})\right)T_c}$$

The following result is then immediately obtained

Theorem 7.3:

1) $\lim_{p\to\infty} \tau(p) = 0$,
2) $\lim_{p\to 1} \tau(p) = 0$,
3) $\tau(p)$ is maximised at

$$p = \frac{\frac{T_c}{T_c+1}}{LambertW\left(-\frac{1}{e}\cdot\frac{T_c}{(T_c+1)}\right) + \frac{T_c}{T_c+1}} \quad ■$$

Remarks 7.2: 1) The behaviour of the aggregate throughput as p goes to its two extremes is as expected. If $p \to 1$ then the nodes do not increase their back-off intervals in response to collisions. The collision probability becomes large and the throughput drops to 0. Obviously, as $p \to \infty$ collisions cause a drastic reduction in attempts essentially shutting the nodes off.

2) In an attempt see what the above asymptotic results have to say about realistic network parameters, in Figure 10 we plot the aggregate throughput for finite K and finite n, using the formula in Eq 7 with equal transmission rate for all the flows. We see that the throughput increases steeply for $1 < p < 2$, but is quite flat with p after $p = 2$. There is an optimal value of p, but unless p is very close to 1, the throughput is not very sensitive to p. It can be seen that the back-off multiplier used

in the standard, i.e., $p = 2$, is adequate unless the number of nodes becomes very large. For $T_c = 17$ (slots), the third part of Theorem 7.3 returns $p = 3.85$, which compares well with the curve for $n = 60$ in Figure 10.

VIII. Application to the Analysis of TCP Controlled File Transfers

A. Some Modeling Assumptions

We will make the following assumptions:

A1: The files are infinitely long. Thus we do not deal with web transfers. Practically, this assumption means that our analysis applies to large file transfers, such as software, document, or media downloads.

A2: The modulation scheme and bit rate of the physical connection between a pair of communicating wireless devices is ideally adapted (but fixed) so that there is no packet loss owing to bit errors. Further, the retransmission time-out at each TCP transmitter is large enough so that time-outs never takes place.

A3: At the transmitter of each wireless device the capacity of the buffer is such that there is no packet loss. This assumption effectively holds in practice if the number of file transfer connections through a node is small enough so that the sum of the maximum TCP windows of all the connections is less than the buffer size. For, say, 10 connections, this would typically require a buffer of no more than 512 KB.

A4: The file transfer throughputs are bottlenecked only by the rates they obtain over the WLAN. For example, the transfers could be between the wireless devices across an ad hoc WLAN, or, in the infrastructure case, between the wireless devices and devices attached to a high speed wired LAN to which the AP is attached. For transfers within a building or campus this assumption is practically valid since most wired LANs are based on 100 Mbps to 1 Gbps Ethernet.

Owing to Assumption A1 it makes sense to talk about the long run time average throughput of a transfer. From Assumptions A2 and A3 it follows that the TCP window of each connection grows to its maximum value, and by Assumption A4, each data packet or ACK of all the TCP connections will be queued at the transmitter of one of the WLAN devices.

Let us adopt the following connection model. There are m connections, indexed by $j, 1 \leq j \leq m$. The source node of connection j is denoted by $s(j)$, and the receiver node is denoted by $r(j)(\neq s(j))$. Thus, for connection j, the TCP ACKs will queue up at the transmitter of node $r(j)$. The data packet length for connection j is denoted by L_j and the ACK packet length by $L_j^{(ack)}$. In general, each node will transmit data packets for some connections and ACK packets for other connections.

In order to use the "saturated queues" analysis presented earlier in the paper, we make the following additional assumption

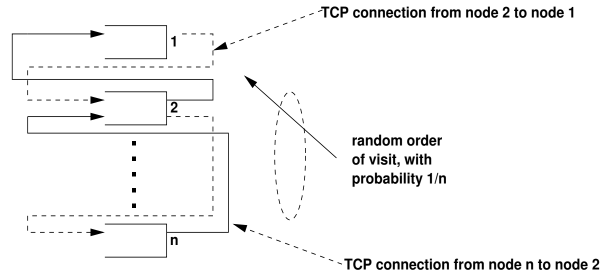

Fig. 11. There are several TCP connections, modeled as "chains" of customers with a fixed population (the window size) circulating in a random polling network. The solid arrows between the queues show the direction of TCP data transfer for a connection, and the dashed arrows show the direction of TCP ACK transmission. The n transmitters are served in random order with equal probability for each node.

A5: The configuration of the TCP connections and the sizes of their windows are such that the transmitter queues of the wireless devices never empty out.

Remark: This assumption is made to permit us to use the fixed point analysis presented earlier in the paper. It, however, considerably restricts the scenarios to which the analysis will apply. For example, the common situation of two or more devices simultaneously downloading files via an AP is not covered by our analysis. This is because the AP needs to send many more packets for each packet that each of the devices sends, and hence the device queues will empty out, violating our saturated queues assumption.

We will utilise Assumption A5 as follows. Recall our discussion in Sec VI. If all the n queues always have packets to send, then they always contend for the channel, and each successful attempt "belongs" to each of the queues with equal probability, $1/n$.

B. A Formula for Connection Throughput

Let us now focus only on the successful attempt instants. Such a success belongs to node i with probability $\frac{1}{n}$. The HOL packet at that node is then transmitted. If this packet is of length L and the transmission rate is C then a time $\frac{L}{C} + T_o$ elapses. If the packet transmitted is a data packet then possibly an ACK is inserted into the transmitter queue of the receiving node (note that if delayed ACKs are used then not every data packet causes an ACK to be generated). On the other hand, if the packet transmitted is an ACK packet then one or more packets are inserted into the transmitter queue of the receiving node. *Thus the queues can be viewed as evolving only at successful polling instants*. This is an important observation as it allows us to ignore the back-off periods while analysing the evolution of the packet queues. Note that this observation does not hold if there are finite rate open-loop arrival processes into the nodes, as these arrival processes will cause the queues to evolve even during back-off periods.

From the above observations, we can now proceed by analysing the discrete time random polling model shown in Figure 11. The discrete "time" in this model evolves over packets. Note that we do not need to be concerned with packet

$$\theta_j(\beta) = \frac{\beta(1-\beta)^{n-1} h_{s(j),j} L_j}{1 + \sum_{i=1}^{n} \beta(1-\beta)^{n-1} \left(\left(\sum_{\{j:s(j)=i\}} h_{i,j} \frac{L_j}{C_{i,r(j)}} + \sum_{\{j:r(j)=i\}} h_{i,j}^{(ack)} \frac{L_j^{(ack)}}{C_{i,s(j)}} \right) + T_o \right) + \left(1 - (1-\beta)^n - n\beta(1-\beta)^{n-1}\right) T_c} \quad (11)$$

lengths (data or ACK), or physical bit rates. We will see that all we need from this model is the fraction of polls to a queue that find packets of each type at the head-of-the-line. There are several TCP connections modeled as "chains" or classes of customers circulating between pairs of nodes. The populations of the chains are the TCP window sizes. If the delayed ACK threshold for a connection is greater than 1 (let us say 2), then at the receiving node for that connection 2 data packets give rise to one ACK packet. We can view this ACK packet as being a batch of 2, that is served together.

The state of the random polling model is the position and type of each packet in each queue. This process evolves over packet times. It is easy to see that the evolution of this rather complicated process is Markovian. Analysis of this Markov chain will yield the following probabilities, that will be used in the throughput formulas.

- $h_{i,j}$: the probability that at a polling instant the HOL packet at node i is a data packet from connection j
- $h_{(i,j)}^{(ack)}$ the probability that at a polling instant the HOL packet at node i is an ACK packet for connection j (for which node i is the receiver node, i.e., $i = r(j)$)

By the observations made just before these definitions, we can conclude that *the probabilities $h_{i,j}$ and $h_{(i,j)}^{(ack)}$ do not depend on data and ACK packet lengths, nor on the physical bit rates of the connections*. These probabilities will depend only on the maximum TCP window sizes, the delayed ACK thresholds, and the connection configurations (i.e., which nodes carry which connections). We also note that once we have these probabilities, the throughput of connection j can be immediately obtained as in Eq 11 (see also Eq 6), where (γ, β) are obtained from the fixed-point analysis. This formula has the same form as the one in Eq 6. In the numerator the term $\beta(1-\beta)^{n-1}$ is the probability that node $s(j)$ has a success, $h_{s(j),j}$ is the probability that the HOL packet belongs to connection j, and when both these events occur connection j has a "reward" of L_j bits. The denominator is the mean length of a back-off and and attempt cycle.

Remarks 8.1:

1) To be technically correct Eq 11 should have been obtained as the ratio of two expectations with respect to the stationary distribution of the Markov chain describing the random polling model. We have shown only the final result in terms of the HOL probabilities at the polling instants, as this is simple and intuitively clear.

2) In Eq 6 the HOL probabilities were obtained from the ratios of the open-loop arrival rates into the queues. In Eq 11, however, the HOL probabilities will need to be obtained from the packet level analysis of the random polling model shown in Figure 11. We will show how this is done in the next subsection.

3) The denominator of the expression now includes a term for the service provided to TCP ACKs.

4) We have used the fact that all data packets within TCP connection j have the same length L_j, and the ACK packets within TCP connection j have the same size $L_j^{(ack)}$. If this were not the case then we would need to make a more elaborate definition of the HOL probabilities which would have to include the probability of finding packets of each possible length.

C. Obtaining the HOL Probabilities

Let λ_j be the throughput of connection j through its sender node $s(j)$ in the random polling model shown in Figure 11. Thus λ_j the average number of packets of connection j that pass through the node $s(j)$ per packet served in the polling model.

Theorem 8.1: If at each success instant one of the nodes is polled with equal probability (i.e., we have the model in Figure 11) then $h_{s(j),j} = \lambda_j n$.

Proof: Let $\pi_{s(j),j}$ denote the fraction of packet services in the model of Figure 11 during which the HOL position at node $s(j)$ is occupied by a data packet of connection j. Since the mean time that a packet spends in the HOL position is n, by Little's Theorem we have

$$\pi_{s(j),j} = \lambda_j n$$

Owing to random polling, the HOL position at node $s(j)$ is observed by a Bernoulli process with probability of "success" equal to $\frac{1}{n}$. Hence by the result that Bernoulli "arrivals" see time averages, we can conclude that

$$h_{s(j),j} = \pi_{s(j),j} = \lambda_j n$$

■

Remarks 8.2: 1) If the throughput of ACKs for connection j through its receiver node $r(j)$ is $\lambda_j^{(ack)}$ then by the same argument as in Theorem 8.1 it follows that $h_{s(j),j}^{(ack)} = \lambda_j^{(ack)} n$.

2) We note that the hypothesis of the Theorem 8.1 that "at each success instant one of the nodes is polled with equal probability" requires the saturation assumption, i.e., Assumption A5, to hold. There are TCP connection configurations for which this assumption will not hold. For example consider a single TCP connection from Node 1 to Node 2. The TCP

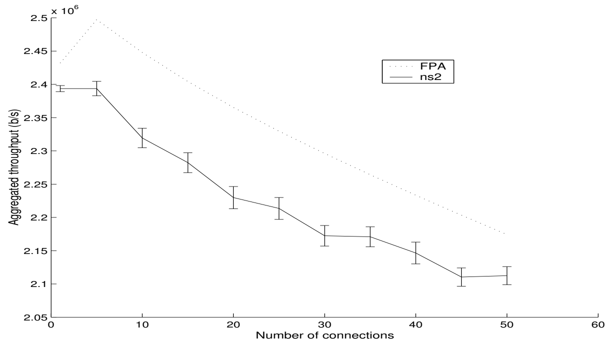

Fig. 12. Plot of aggregate TCP controlled file transfer throughput vs. number of simultaneous transfers over an IEEE 802.11b network. Results obtained from the approximate analysis in the paper and from *ns2* simulations are shown. We show 95% confidence intervals around the simulation results. In the *ns2* simulation the IEEE 802.11 parameters are used; the data rate is 11 Mbps and the control rate (used to send RTS, CTS and ACK) is 2 Mbps.

receiver uses a delayed ACK threshold of 2; i.e., it returns one ACK for two received data packets. Clearly over a large number of packet transmitted we cannot say that about half will come from Node 1 and the other half from Node 2. In this case the receiver node will tend to empty out and the saturation assumption will not apply. On the other hand if Node 2 was also sending to Node 1 then our analysis will apply.

3) In view of Theorem 8.1 we need to analyse the random polling model and obtain the λ_js and the $\lambda_j^{(ack)}$s, and this will yield the HOL probabilities needed in the throughput formula.

D. *Comparison with* ns2 *Simulations*

In Figure 12 we compare the results from the analysis presented above and *ns2* simulations; 95% confidence intervals are shown around the simulation results. For comments on the version of *ns2* used, see Sec V-A. The scenario simulated is that there are n nodes paired up with n other nodes; each node in the first group is performing a TCP controlled long file transfer to its corresponding member in the other group. The maximum receiver window for each TCP connection is 20 packets, the TCP packet length is 1 KB, and the receivers do not delay the ACKs (i.e., an ACK is returned for each received data packet). In this situation, of course, $h_{i,j}$ will be 1 whenever Node i is the source node of connection j, and $h_{i,j}^{ack}$ will be 1 whenever Node i is the receiver node for connection j. The physical link rates are all 11 Mbps. The aggregate throughput over all the connections is plotted vs. the number of connections. We notice that the match between analysis and simulations is good, with the worst case error being about 6%. The simulation trace file showed that during the simulations there was no TCP time-out; thus our Assumption A2 held in this case.

Another scenario that we evaluated was two nodes sending files to each other, simultaneously. In this case the aggregate throughput predicted by the model is 2.4164 Mb/s, while *ns2* simulations return a 95% confidence interval of [2.4054,2.4153] Mb/s. In the simulation, the throughput obtained by each transfer is approximately a half of the aggregate throughput.

IX. SUMMARY

Our analysis has provided a simple and general representation of the fixed point equation that arises from an analysis initiated by Bianchi in [2]. The representation is insensitive to the distribution of the back-off times. We show that if the mean back-off durations for successive retrials are monotone nondecreasing then the fixed point equation has a unique solution. Then we provide general throughput formulas for open-loop arrival processes (e.g., UDP transfers). We recover the observation that connections with small physical rates dominate the throughputs of other connections. We then turn to the special case of exponential back-off with an arbitrary positive multiplier, p, and where we do not limit the number of retrials a node can make. This leads to simpler expressions which permit us to study the network performance as the number of nodes goes to infinity. For this case, we obtain a characterisation of the fixed point solution for the collision probability for each n. Then we take n to ∞ and obtain the limit of the collision probability and aggregate attempt rate that agree with the results of Kwak et al in [6]. We also provide a relaxed fixed point iteration for computing the fixed point for any finite n when the number of retrials is not limited. The asymptotic aggregate throughput is obtained and from this the optimal back-off multiplier p is also derived.

For exponential back-off, and geometrically distributed back-off periods, the back-off process can be modeled via a discrete time Markov chain. In Section V-B we study this DTMC, and for some simple computable cases we compare the collision probability obtained from the DTMC with that obtained from the fixed point analysis.

Finally, we show how the saturation throughput analysis can be used to obtain TCP controlled file transfer throughputs for some network scenarios. In this analysis we exploited the idea that, for window controlled traffic, the back-off process evolution can be decoupled from the packet service process, the latter being modeled by a random polling queue.

APPENDIX

Proof: (Lemma 5.1) We have

$$G(\gamma) \quad := \quad \frac{1+\gamma+\gamma^2\cdots+\gamma^K}{b_0+\gamma b_1+\gamma^2 b_2+\cdots+\gamma^k b_k+\cdots+\gamma^K b_K}$$

and we need to show that the derivative of this function with respect to γ is negative. Taking the derivative we find that we need to show that

$$\sum_{k=0}^{K} b_k\gamma^k\left(\sum_{j=1}^{K} j\gamma^{(j-1)}\right) \leq \sum_{k=0}^{K}\gamma^k\left(\sum_{j=1}^{K} jb_j\gamma^{(j-1)}\right)$$

i.e.,

$$\sum_{k=0}^{K}\sum_{j=1}^{K} jb_k\gamma^{(k+j-1)} \leq \sum_{k=0}^{K}\sum_{j=1}^{K} jb_j\gamma^{(k+j-1)}$$

or, equivalently, we need to show that

$$\sum_{n=1}^{2K} \gamma^{(n-1)} \sum_{\substack{j=\max\{(n-K),1\} \\ k=(n-j)}}^{\min\{n,K\}} j(b_j - b_k) \geq 0$$

Now we consider each term $\sum_{\substack{j=\max\{(n-K),1\} \\ k=(n-j)}}^{\min\{n,K\}} j(b_j - b_k)$ and show that it is nonnegative. To this end, define

$$m(n) = |\{(j,k) : j + k = n, 1 \leq j \leq K, 0 \leq k \leq K\}|,$$

where $|\cdot|$ denotes set cardinality. When $k = j$, $jb_j - jb_k = 0$ and the corresponding term vanishes from the sum. Also, k equals 0 only when $j = n$ and $1 \leq n \leq K$. Hence, simplifying the above expression, we get,

$$\sum_{j=\max\{(n-K),1\}}^{\max\{(n-K),1\}+\lfloor \frac{m}{2} \rfloor - 1} ((n-j)-j)\,(b_{n-j} - b_j) + n(b_n - b_0)1_{\{1 \leq n \leq K\}}$$

which is nonnegative since, in the range of the sum, $(n - j) - j \geq 0$ and $b_{n-j} - b_j \geq 0$. It is also easily seen that the derivative of $G(\cdot)$ is strictly negative for $\gamma > 0$ if the b_k are not all equal, this implies that $G(\cdot)$ is strictly decreasing in this case. ■

Proof: (Lemma 7.1)

1) For $x \geq 0$, write $z(x) = LambertW(axe^x)$, i.e.,

$$z(x)e^{z(x)} = axe^x$$

It is easily seen that for $x > 0$, $z(x) > 0$, and $z(x) \uparrow \infty$ for $x \to \infty$. Now, taking natural logarithms, we obtain, for all $x > 0$,

$$\ln z(x) + z(x) = \ln ax + x$$

or

$$\frac{z(x)}{x} = \frac{\frac{\ln ax}{x} + 1}{\frac{\ln z(x)}{z(x)} + 1}$$

which, on taking $x \to \infty$ yields the desired result since $\frac{\ln ax}{x}$ and $\frac{\ln z(x)}{z(x)}$ both go to 0.

2) By definition, $LambertW(xe^x) = x$, and LambertW is monotone increasing for positive arguments. Hence, for $0 < a \leq 1$, $LambertW(axe^x) \leq x$.

3) Follows by combining the previous two parts. ■

Acknowledgements

We are grateful to Chadi Barakat (INRIA, Sophia Antipolis, France) for sparing his time for some useful discussions.

References

[1] G. Berger-Sabbatel, F. Rousseau, M. Heusse, and A. Duda. Performance anomaly of 802.11b. In *Proceedings of IEEE Infocom 2003*. IEEE, 2003.

[2] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.

[3] F. Cali, M. Conti, and E. Gregori. IEEE 802.11 protocol: Design and performance evaluation of an adaptive backoff mechanism. *IEEE Journal on Selected Areas in Communications*, 18(9):1774–1780, September 2000.

[4] Anurag Kumar and Deepak Patil. Stability and throughput analysis of unslotted cdma-aloha with finite numer of users and code sharing. *Telecommunication Systems*, 8:257–275, 1997.

[5] Anurag Kumar, Eitan Altman, Daniele Miorandi, and Munish Goyal. New insights from a fixed point analysis of single cell IEEE 802.11 wireless LANs. Technical Report RR-5218, INRIA, Sophia-Antipolis, France, June 2004.

[6] B.-J. Kwak, N.-O. Song, and L. E. Miller. Analysis of the stability and performance of exponential backoff. In *Proceedings of IEEE WCNC*, 2003.

[7] Stefan Mangold, Sunghyun Choi, Peter May, Ole Klein, Guido Hiertz, and Lothar Stibor. IEEE 802.11e wireless LAN for quality of service. In *Proc. European Wireless (EW 2002)*, February 2002.

[8] Venkatesh Ramaiyan, Anurag Kumar, and Eitan Altman. Fixed Point Analysis of Single Cell IEEE 802.11e WLANs: Uniqueness, Multistability, and Throughput Differentiation. Submitted, November 2004.