Improving Fairness in Multi-Hop Mesh Networks Using 802.11e

K. Duffy, D.J. Leith, T. Li and D. Malone
Hamilton Institute, NUI Maynooth

Abstract— In this paper we introduce a tractable analytic model of throughput performance for general 802.11 multi-hop networks. We use this model to explore a number of fairness issues in 802.11 multi-hop networks that have a significant impact on performance and capacity for realistic traffic. Schemes using the functionality provided by 802.11e are proposed to mitigate this unfairness and their efficacy demonstrated.

I. INTRODUCTION

IEEE 802.11 technology has been enormously successful, with wireless 802.11a/b/g edge networks becoming increasingly pervasive. Much of the uptake of this technology has been confined to single-hop infrastructure mode situations, for example in offices and DSL/wireless home networks. Multi-hop mesh networks based on 802.11 technology do exist, often in the context of community WLANs, but their design and performance remain poorly understood. The analysis and design of flexible, efficient and simple-to-use WLANs with more than one hop remains a challenging problem.

Building on our previous work in [1], [2], in this paper we introduce a tractable analytic model of throughput performance for general 802.11 multi-hop networks. To our knowledge this is the first multi-hop analytic 802.11 model that supports both finite loads and general multi-radio multi-channel network topologies. We argue that consideration of finite load is essential in a mesh network context as (i) relaying of traffic generally carries a contention and/or packet loss overhead at each hop. Even if stations at the first stage in a relay are saturated, stations at subsequent stages are not. Modeling the scaling behavior of throughput with number of relay stages and with relay topology therefore makes consideration of finite loads essential. (ii) Realistic traffic such as voice and web is low-rate and on/off. Network performance with such traffic cannot be accurately modeled without consideration of finite loads.

We use this model to explore a number of fairness issues that arise at aggregation points in 802.11 multi-hop networks and that have a considerable impact on network performance and capacity. These fairness issues are a feature of the 802.11 CSMA/CS contention mechanism and are quite different in nature from the types of fairness issue previously discussed in the context of general multi-hop wireless networks.
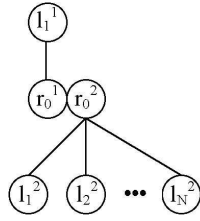
The following example illustrates one important form of 802.11 unfairness. Consider the simple 802.11b multi-hop network depicted in Figure 1(a). Here, wireless station l_1^1 has a wired back-haul connection and communicates with the wireless clients $l_1^2, ... l_N^2$ via the wireless relay station r_0^1/r_0^2.

The latter denotes a single relay station with two radios. Node superscripts indicate the channel used by a node. Nodes l_1^1, r_0^1 operate on the same radio channel, orthogonal to that used by nodes $l_i^2, i = 1, ..., N$, and relay node r_0^2. Suppose now that the network carries two-way voice calls between client stations $l_i^2, i = 1, ..., N$ and back-haul gateway l_1^1. Voice calls are modeled as on-off 64Kbs traffic[1]. Figure 1(b) shows the throughput as the number of active voice calls is increased. We can see that when more than eight voice calls are active, the throughput of the downstream calls begins to fall even though the upstream throughput continues to increase. It is the throttling of the downstream halves of the voice calls, rather than the physical radio bandwidth, that limits the network voice call capacity as a call will be dropped once the loss rate of one half of the conversation becomes too great. This behavior occurs because in 802.11 networks the MAC layer contention mechanism allocates a roughly equal share of transmission opportunities to every wireless station. That is, the client stations $l_i^2, i = 1, ..., N$ have roughly the same number of transmission opportunities as the relay station r_0^2. However, the relay station is required to transmit the downstream part of N voice calls whereas each client station only transmits the upstream part of a single voice call. As the number of calls is increased, the relay station is unable to win sufficient transmission opportunities to support all downstream calls. Although we consider a simple multi-hop topology in this example, it is evident that this type of behaviour can be expected to be common in general: all that is needed is the presence of one or more relay stages where flow aggregation takes place.
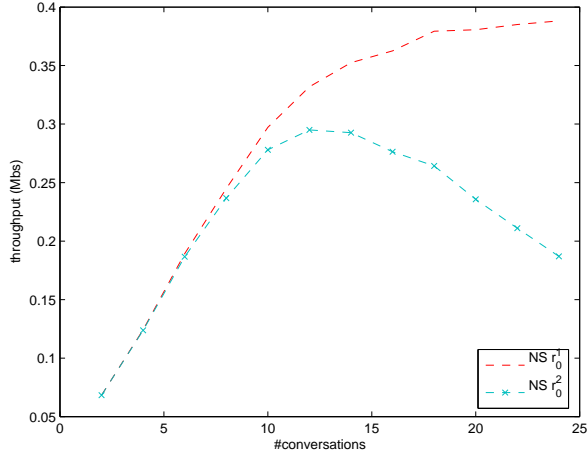
In summary, the contribution of this paper includes the following:

1) The derivation of a finite-load Markov chain model of CSMA/CA throughput in general multi-hop/mesh network topologies. To our knowledge this is both the first finite-load mesh model and the first predictive CSMA/CA model for general multi-hop network topologies.

2) The demonstration of upstream/downstream unfairness induced by the 802.11 MAC at aggregation points in a relay backbone. We show that, in the context of voice traffic, it is this unfairness that limits network capacity

---

[1]Parameters for the voice calls are taken from [3]: 64kbps on-off traffic streams where the on and off periods are distributed with mean 1.5 seconds. Traffic is two-way; the on period of an upstream call corresponds to the off period of its downstream reply.

(a) Network topology

	Duration(μs)
Slot time, σ	20
Propagation delay, δ	1
$CW_{min} = 32\ \sigma$	640
DIFS (AIFS=0)	50
SIFS (Short Inter Frame Space)	10
PLCP Header @1Mbps	192
MAC Header 24 Bytes @1Mbps	17.5
CRC Header 4 Bytes @1Mbps	2.9
IP Header 20 Bytes @11Mbps	14.5
MAC ACK 14 Bytes @11Mbps	10
E_i payload 80 Bytes @11Mbps	58.2

TABLE I

802.11B MAC VALUES USED THROUGHOUT THE EXAMPLES IN THIS
PAPER.

(b) Network throughput

Fig. 1. Example of aggregate throughput vs number of voice calls for the multi-hop 802.11b WLAN topology in (a). Voice packets are transported between l_1^1 and $l_1^2, ..., l_N^2$ by node r_0^1/r_0^2 which denotes a relay station with two radios. Stations with superscript 1 use an orthogonal radio channel to stations with superscript 2 to avoid interference. NS simulation, 802.11 MAC parameters in Table I.

rather than radio bandwidth. This issue was known to be important in infrastructure mode networks [4], [5] and we demonstrate that it is also important in mesh networks.

3) We propose and analyse a simple scheme that uses the flexibility provided by the new 802.11e standard (specifically, TXOP and CWmin adjustment) to restore fairness at relay aggregation points. The proposed scheme is applicable to any mesh network with loop-free routing.

4) We demonstrate an unfair behavior induced by the 802.11 MAC at aggregation points due to load imbalances. We analyse the cumulative impact of this unfairness over a multi-hop relay and briefly demonstrate the feasibility of using 802.11e functionality to restore fairness, leaving more detailed consideration of this as future work.

II. RELATED WORK

The analysis and design of multi-hop wireless networks has been the subject of considerable interest. There exists a considerable body of work on the general problems of routing, channel allocation, power management, scheduling and flow control. Much of this work is either in the context of general mathematical optimization frameworks and/or relates to "clean slate" design. With regard to the literature concerned specifically with 802.11-based multi-hop networks, much of the published work has focused on issues such as routing and interference management (e.g. see [6], [7], [8], [9] and references therein) and changes to the 802.11 MAC to enhance performance (e.g. see [10], [11], [12] and references therein). Analytic modeling of the throughput performance of the 802.11 CSMA/CA MAC in a multi-hop context has received relatively little attention. In [13] the feasibility of routing traffic through a multi-hop Aloha network using a model with certain saturation assumptions. Sun *et al.* [14] consider the use of a single-hop saturated throughput model to support adaptive routing in multi-hop networks. Gao *et al.* [15] focus on the saturated modeling of hidden node behavior in path and grid topologies. Wang *et al.* [16] consider a simplified throughput model in a random Poisson topology with saturated nodes. Barowski *et al.* [17], [18] consider a finite load model of 802.11 in the context of single hop networks; the authors comment that the model is only valid for light loads (where achieved throughput is close to the offered load). The authors briefly mention that this work can be extended to multi-hop networks by specifying the offered load on a node and including a complex queueing component in the model but the resulting model is not predictive (the offered relay load needs to be pre-specified at each node) and no analysis is presented.

III. MODELING

A. General setup

We consider a general 802.11 mesh network where some stations only have the capacity to communicate locally on a single frequency and there are relay stations with multiple radios to relay data to other channels. We set the network model up without explicit reference to the network's topology. The topology enters by limiting the routes along which data

can be relayed. We say that groups of stations that can communicate on a single frequence are in a "zone". We assume that stations have been allocated to zones so that within a particular zone there are no hidden stations and collisions only occur when more than one station attempts to use the medium. This requires sufficient available channels and a suitable channel allocation scheme such as [19].

Consider a wireless network with M distinct wireless zones. Each zone talking on a frequency that does not overlap with its neighbors. In earlier work [1], [2], we considered a single infrastructure mode network consisting of wireless stations employing 802.11's access protocol, assuming it has no hidden nodes and collisions are only caused by more than one station within a zone attempting to transmit simultaneously. We developed a finite-load model based on a mean-field Markov chain approximation that was shown to accurately capture and predicted key features of network performance. We shall employ that model within each network zone to determine that zone's performance, but each zone's performance is coupled to other zones because of relayed traffic.

In particular, in the network we assume there are two sorts of stations: *local stations* that generate traffic and transmit within a single zone; and *relay stations* that have multiple radios, do not generate traffic themselves, but relay traffic from local or relay stations in other zones into their own zone.

For each zone $n \in \{1, \ldots, M\}$ we label local stations as elements of the set $\mathcal{L}_n = \{l_1^n, \ldots\}$ and relay stations as elements of the set $\mathcal{R}_n = \{r_0^n, \ldots\}$, whilst allowing $\mathcal{L}_n$ or $\mathcal{R}_n$ to be the empty set. Within each zone we employ our model which consists of an embedded Markov chain; each zone's Markov chain runs in a different time frame. We assume that all local nodes have defined stochastic arrival rates corresponding to the exogenous load on the station. That is, for each $n \in \{1, \ldots, M\}$, $l \in \mathcal{L}_n$ has a defined probability of packet arrival q_l during each transition in the Markov chain; later we will show how to relate q_l to offered load in bits/sec, to make the model predictive. For each relay station, $r \in \mathcal{R}_n$ where $n \in \{1, \ldots, M\}$, we must determine the probability it is offered a packet, q_r, as a function of routing information and network zone throughput performance.

First consider a fixed zone $n \in \{1, \ldots, M\}$. Let $p_c \in [0, 1]$ denote the probability of collision given attempted transmission for station $c \in \mathcal{R}_n \cup \mathcal{L}_n$ and q_c denote the probability a packet arrives during a state transition. Let $\tau_c \in [0, 1]$ denote the stationary probability that station c transmits in a given slot. For given (p, q), $\tau = \tau(p, q)$ is given in [1], [2] by

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

where the normalization constant η is

$$
\begin{aligned}
\eta =\ & \frac{qW_0}{1-(1-q)^{W_0}} + \frac{qW_0(qW_0 + 3q - 2)}{2(1-q)(1-(1-q)^{W_0})} \\
& + (1-q) + \frac{q(W_0+1)(p(1-q)-q(1-p)^2)}{2(1-q)} \\
& + \frac{pq^2}{(1-q)2(1-p)} \left(\frac{W_0}{1-(1-q)^{W_0}} - (1-p)^2 \right) \\
& \left(\frac{2W_0(1-p-p(2p)^{M-1})}{(1-2p)} + 1 \right)
\end{aligned}
\quad (2)
$$

and W_0 is the station c's minimum contention window. If q_r was known a priori for each $r \in \mathcal{R}_n$, then p_c for each $c \in \mathcal{R}_n \cup \mathcal{L}_n$ would be completely determined by solving the following of non-linear equations. They state that the probability station c does not experience a collision given it is attempting transmission is the probability that no other station within its zone is attempting transmission:

$$1 - p_c = \prod_{b \in \mathcal{R}_n \cup \mathcal{L}_n,\ b \neq c} (1 - \tau_b), \quad (3)$$

and we substitute $\tau_b = \tau(p_b, q_b)$ from Equation 1.

We must relate the behavior of distinct zones in order to connect the transmissions in one zone to the arrivals in the next. We first note that for each zone we have distinct embedded Markov chain systems and thus each chain system runs in its own time frame. In terms of τ, the timescale E_n is

$$E_n = \left(\prod_{b \in \mathcal{R}_n \cup \mathcal{L}_n} (1 - \tau_b) \right) \sigma + L(1 - \prod_{b \in \mathcal{R}_n \cup \mathcal{L}_n} (1 - \tau_b)),$$

where each packet takes L seconds to be transmitted on the medium and the idle slot-length is σ seconds[2]. For convenience, we define a function that for any station has the value of the station's zone's expected transition time. That is, for all $m \in \{1, \ldots, M\}$ and each station $c \in \mathcal{R}_m \cup \mathcal{L}_m$, we define $E(c) = E_m$.

However the parameter q_r is not known a priori for each relay station, so we must now determine the offered load from a relay station into its zone. We start by defining for each local station $l \in \mathcal{L}_n$, $n \in \{1, \ldots, M\}$, a fixed route f_l from its zone to a destination zone. Its route is defined by an ordered set of relay stations through which l's packets must pass and a local station which is in the traffic's destination zone. While we do not assume that each zone appears less than twice in the route, we do insist that no relay station appears more than once. The route is defined by

$$f_l = \{l, s_1 \ldots, s_m, d\},$$

where d, a local station in the destination zone, is in the same zone as s_m. If $m = 0$, then l and d are in the same zone and no relaying is necessary. While modeling throughput, it is not actually important that we choose a specific destination as all stations within a zone hear all local transmissions. We assume routes are pre-determined by an appropriate wireless routing protocol.

Now, for each relay r, we will derive an expression for q_r, in terms of a sum over paths that use r. Let q_{l,s_k} denote the probability that a packet from l is offered to s_k. We assume the proportion of traffic from l that makes it to s_k is the part of s_{k-1}'s throughput in the proportion $q_{l,s_{k-1}}/\gamma_{s_{k-1}}$, where $\gamma_{s_{k-1}}$ represents the total traffic arriving s_{k-1}, defined below in Equation (4). In particular, the probability l transmits

---

[2]This time, L, includes DIFS and so forth and E_n includes time lost due to MAC level retransmissions. We will slightly simplify the model by assuming that transmissions and collisions use the same time on the medium, but this assumption is easily relaxed.

successfully during a period $E(l)$ is the probability it attempts transmission and does not experience a collision: $\tau_l(1-p_l)$. To determine the likelihood a packet is offered to s_1 by l during a period of length $E(s_1)$, we rescale to take into account the fact that the average time for a counter decrement in each zone is different: $q_{l,s_1} := \gamma_{s_1}^{-1}(\tau_l(1-p_l))E(s_1)/E(l)$. Thus for $k \in \{1, \ldots, m\}$:

$$q_{l,s_k} = \left(\prod_{i \leq k} \frac{1}{\gamma_{s_i}}\right) \frac{E(s_k)}{E(l)} \prod_{b \in f_l, b < s_k} \tau_b(1-p_b),$$

where $b < s_k$ implies b comes before s_k in the ordered list f_l. The final objects we need to define are γ_r and q_r for each relay r. These are given by the following:

$$\gamma_r := \sum_{P_r} q_{l,s} \frac{\tau_s(1-p_s)E(r)}{E(s)}, \quad q_r = \min\{1, \gamma_r\} \quad (4)$$

where the sum is over all elements of P_r, the collection of routing paths in which r is a relay station and s immediately precedes r, and for each element of the sum l is the local station where the data is first transmitted.

It now possible, for given inputs q_l, to numerically solve the model using the equations (3) by searching[3] for q_r that satisfy equation (4). The throughput of station l to its destination is given by the probability a packet from l arrives in d's zone times the packet payload P bits:

$$P\left(\prod_{i=1}^{k} \frac{1}{q_{s_i}}\right) \frac{1}{E(l)} \prod_{b \in f_l, b < d} \tau_b(1-p_b), \quad (5)$$

Thus for given inputs q_l, the model is completely solved. In the following example we demonstrate how to relate the inputs q_l to offered load in bits per second.

B. Example

Consider the two-channel two-hop network depicted in Figure 1(a). We have $\mathcal{R}_1 = \{r_0^1\}$, $\mathcal{L}_1 = \{l_1^1\}$, $\mathcal{R}_2 = \{r_0^2\}$ and $\mathcal{L}_2 = \{l_1^2, \ldots, l_N^2\}$. The local stations in $\mathcal{L}_1$ and $\mathcal{L}_2$ communicate via the relays in $\mathcal{R}_1$ and $\mathcal{R}_2$. Thus we define the routes $f_{l_1^1} = \{l_1^1, r_1^1, l_1^2\}$ and, for each $n \in \{1, \ldots, N\}$, $f_{l_n^2} = \{l_n^2, r_1^2, l_1^1\}$. Note that any elements of zone 2 can be the recipient of l_1^1's traffic and that, although we have logically separated r_1^1 and r_1^2, they could be the same physical device and only one station with two radios is necessary for this scenario.

To model $64Kbs$ two-way voice conversations, we must solve the model looking for $q_{l_1^1}$ and $q_{l_n^2}$, $n \in \{1, \ldots, N\}$, that correspond to offered loads of $N \times 32Kbps$ and $32Kbps$, respectively. With 80 byte packets, this corresponds to 50 packets per second for l_n^2 and $50 \times N$ packets per second for l_1^1. In order to move between model and simulation arrival rates, we use the following logic[4]. Since we have small buffers

---

[3]This is analogous to searching for a fixed-point in Bianchi based models. We do not consider the existence or uniqueness of such solutions here.

[4]The validity of this approach is discussed in more detail in [1], [2], albeit in the context of single-hop infrastructure mode 802.11 networks, and is demonstrated in Figure 2. It is also possible to readily accommodate long buffers using a similar approach when the traffic is Poisson or saturated.
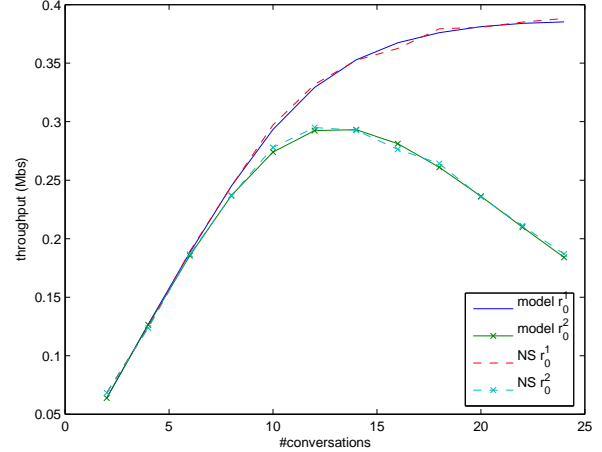

Fig. 2. Example of Figure 1 comparing NS packet-level simulation results and model predictions. Network topology is shown in Figure 1(a), NS simulation parameters as in Figure 1.

(to minimise latency as voice traffic is delay-sensitive), the parameter q is the probability that at least one packet arrives in the expected time spent per state E. In simulation, the probability that at least one packet arrives during E is one minus the probability that the first inter-packet time is greater than E. Hence, when inter-packet arrival times are exponentially distributed we have for each $l \in \mathcal{L}_1$, $q_l = 1 - \exp(-\lambda_2 E(l))$ with rate $\lambda_2 = 1/50$ and $q_{l_1^1} = 1 - \exp(-\lambda_1 E(l_1^1))$ with $\lambda_1 = 1/(50N)$.

The quantities of interest are then the throughputs of the stations l_1^1 and l_n^2, given by equation (5). Figure 2 compares the resulting model predictions with NS packet-level simulation results. It can be seen that the model is remarkably accurate.

C. Model scope

For ease of presentation, the model assumes a fixed packet size. This assumption can be relaxed, with the primary difficulty being that each zone's Markov chain transition time depends on the distribution of packet sizes. Channel errors are assumed to be caused only by collisions. To a first order of approximation, random noise can be introduced by independently dropping packets, as 802.11 treats this in the same way as collisions. We have considered a simple queue model in this paper, however it is clearly possible to introduce more complex traffic and queue models without adding further states into the Markov chain. For example, the q values could be calculated using more elaborate queueing modeling. Also, the probability that a station's buffer is empty immediately after successful transmission could be made dependent on the backoff stage at which that transmission took place. These probabilities could be obtained from traffic and queue modeling or even estimated from a running system. Alternatively, larger buffers could be explicitly modeled by significantly increasing the number of states in the Markov chain.
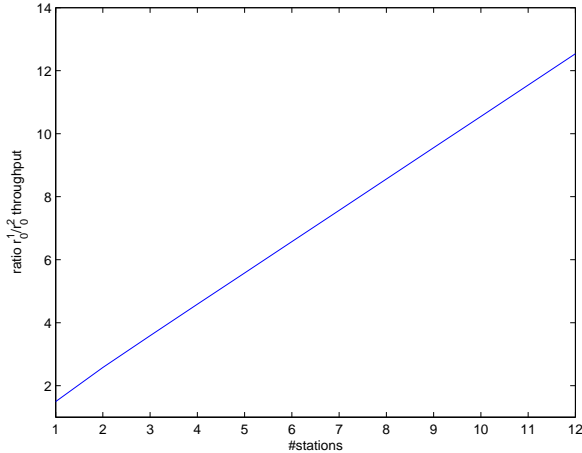
Fig. 3. Ratio of throughputs of competing greedy client and relay UDP streams vs number of streams (two-hop topology in Figure 1).

Fig. 4. Example of Figure 1, now with relay stations prioritised using $TXOP$ scheme. It can be seen that fairness is restored between the upstream and downstream parts of the two-way voice calls. Network topology is shown in Figure 1(a).

IV. RESTORING FAIRNESS AT RELAY AGGREGATION POINTS

The 802.11 CSMA/CA mechanism provides stations sharing the same zone (i.e. in transmission range and sharing the same radio channel) with approximately the same number of transmission opportunities. This includes not only the client stations, but also any relay stations. Suppose we have n_u client stations and n_r relay stations. Let n_{d_i} denote the number of clients for which relay station i is carrying relay traffic. The n_u client stations have roughly a $n_u/(n_u + n_r)$ share of the bandwidth while relay station i has only a $1/(n_u + n_r)$ share despite carrying n_{d_i} flows. It is this asymmetry that can result in the relay becoming the network bottleneck.

The validity of this argument, at least for greedy (every station always has a packet to send) flows, can be seen from Figure 3. The figure shows the ratio of the throughputs achieved in the two-hop topology shown in Figure 1 by competing client and relay UDP flows as the number of flows is varied (with a single relay station and an equal number of client stations $n_u = N$ and relay streams $n_d = N$). The throughput ratio is linear and equal to the number n_u of client stations.

While the foregoing argument provides insight and makes accurate predictions for greedy traffic flows, the situation with voice calls is more complex. We can see this immediately from Figure 1 where the upstream and downstream voice flows achieve almost equal throughput up to around 8 calls. In contrast, if the upstream/downstream flows were greedy (always have a packet to send) then the foregoing analysis indicates that with 8 calls the upstream flows would in aggregate achieve a factor of 8 greater throughput than the downstream flows. We can understand this behavior by noting that, firstly, voice traffic is relatively low rate and so need not make use of every available transmission opportunity awarded by the 802.11 MAC. Secondly, a voice conversation involves
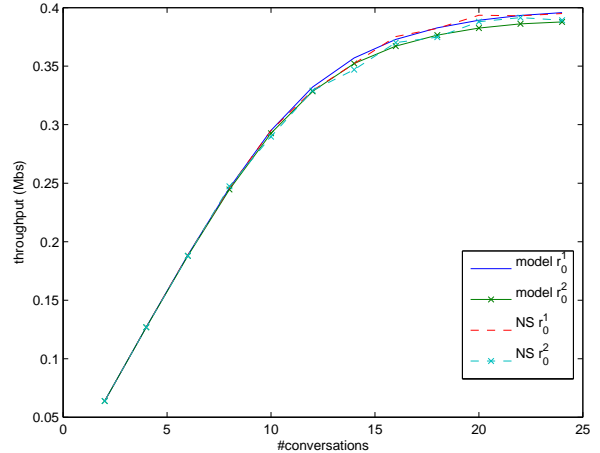
speakers approximately taking turns talking. That is, traffic is between pairs of speakers with the on period of one speaker roughly corresponding to the off period of the other. Both of these features mitigate the contention between the client and relay stations for access to the wireless channel.

Again, consider a wireless zone with n_u client stations and n_r relay stations. Let n_{d_i} denote the number of clients for which relay station i is carrying relay traffic[5]. We consider the fairness requirement that each client station has an equal share of transmission opportunities, but our approach can be readily generalized to other fairness measures. We propose that the $TXOP$ packet bursting mechanism in 802.11e provides a straightforward and fine grained achieving fairness. Specifically, we simply assign a $TXOP$ value of one packet to each client station and a $TXOP$ value of n_{d_i} to relay station i. Since each relay station gains a $1/(n_u + n_r)$ share of transmission opportunities, by transmitting n_{d_i} packets (one packet from each of the n_{d_i} streams) at each transmission opportunity it can be immediately seen that we restore the $n_{d_i}/(n_u + \sum_{i=1}^{n_r} n_{d_i})$ fair share to the relay traffic.

The effectiveness of this scheme is shown in Figure 4, where it can be seen to restore fairness between the upstream and downstream voice calls in our previous example. A second example with a more complex topology is shown in Figures 5 and 6. It is interesting to observe that with standard $TXOP$ settings it is the throughput of the downstream part of the voice calls, which relies upon the relay station for forwarding, that first falls below 90% of the offered load (marked by the dashed line in the figure) at which point the quality of the voice call is likely to have deteriorated to the point that the call would

---

[5]This can be implemented in practice by inspecting the forwarding interface queue. This provides a direct measure of the number of active relay streams in a manner which is both straightforward and dynamically adapts to accommodate bursty and intermittent traffic.
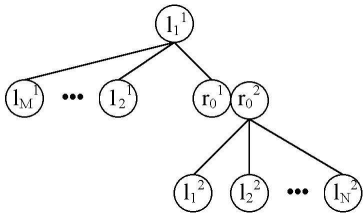
Fig. 5. Example multi-hop network topology. Stations l_i^1, r_0^1 with superscript 1 use an orthogonal radio channel from stations l_i^2, r_0^2 with superscript 2 to avoid interference, with node r_0^1/r_0^2 denoting a two-radio relay station.

be dropped. This occurs with only eight voice calls, and it is the unfairness between the upstream and downstream halves of the two-way voice calls that limits network capacity rather than the available radio bandwidth. With the proposed $TXOP$ scheme it can be seen that fairness is essentially restored and, moreover, the capacity rises to 14 voice calls before throughput falls to 90% of the offered load, an increase of 55% in voice call capacity.

We comment that with this $TXOP$ approach a relay station transmits n packets in a single burst. For n large, this can result in the station occupying the channel for a substantial consolidated period of time and this may, for example, negatively impact competing delay sensitive traffic. We can address this issue in a straightforward manner by using multiple smaller bursts instead of a single burst. When using smaller packet bursts, it is necessary to ensure a corresponding increase in the number of transmission opportunities won by the station. This can be achieved by using a smaller value of CW_{min} for the prioritised traffic class at the station. It is shown in [20] that competing traffic classes gain transmission opportunities approximately in inverse proportion to their values of CW_{min}. Let k denote the ratio of the stations upstream class CW_{min} value to that of the prioritised class at the station. Scaling k with the number of transmission opportunities required provides coarse (recall that in 802.11e k is constrained to be a power of two) prioritization of downstream flows. We then complement this with use of $TXOP$ for fine grained adjustment of the packet burst lengths, scaling $TXOP$ with $1/k$. Hence fine grained prioritization can be achieved while avoiding unduly large packet bursts.

V. UNFAIRNESS DUE TO DIFFERENCES IN LOAD

In this section we use the model to consider the impact of load imbalances on the throughput fairness at relay aggregation points. Consider a single 802.11 hop with one client station transmitting a 64Kbs voice call and one competing client station. Figure 7 plots the achieved throughput of the voice station as the offered load on the competing client station is increased. Results are also shown for greater numbers of competing stations, each having the offered load shown on the x-axis. It can be seen that as their offered load is increased the competing stations are able to steal bandwidth from the voice call. The throughput of the voice station can be much less than the fair share that it could achieve if it had a higher offered
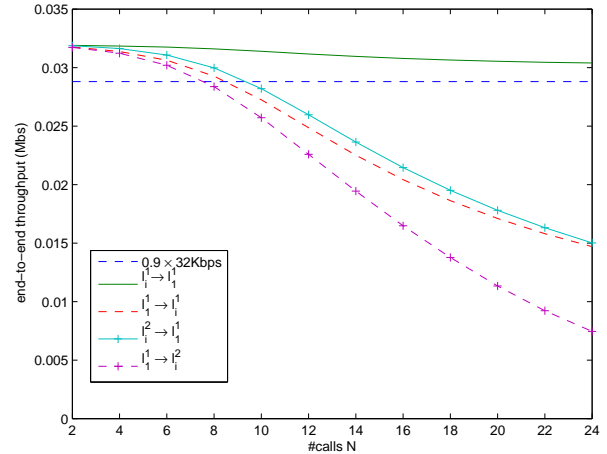
(a) with standard 802.11b $TXOP$ settings.
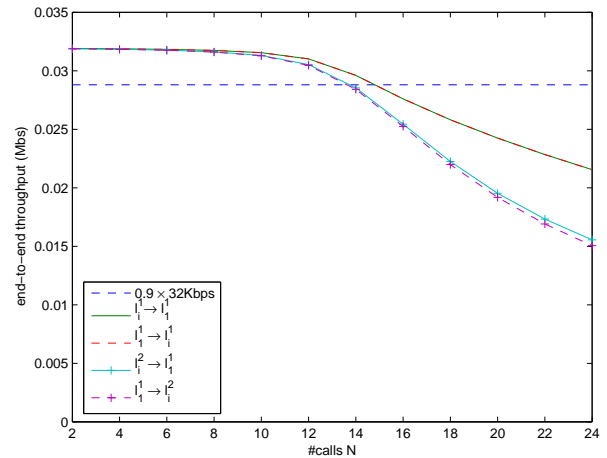
(b) with $TXOP$ prioritisation.

Fig. 6. End-to-end throughput of two-way voice calls in multi-hop network with tree topology shown in Figure 5 and using $M = 2$ and the value of N marked on the x-axis. Voice calls take place between $l_2^1, ..., l_M^2, l_1^2, .., l_N^2$ and l_1^2, with r_0^1/r_0^2 relaying calls between l_1^1 and $l_1^2, ..., l_N^2$. Upper plot (a) shows the throughput when standard 802.11b $TXOP$ settings are used (1 packet per transmission opportunity) and the number N of l_i^2 stations is varied. Model results. Lower plot (b) shows throughput with our $TXOP$ prioritisation scheme.
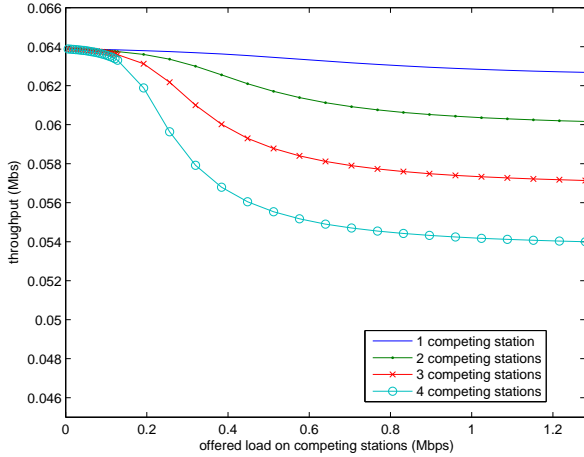
Fig. 7. Impact on achieved throughput of differences in offered load. Single hop with one station transmitting a 64Kbs stream and a varying number of competing stations. Model results. All competing stations have the same load, marked on the x-axis.

(a) Voice call throughput

(b) Throughput of competing stations.

Fig. 8. Cumulative impact over multiple hops of throughput unfairness in 802.11 between stations with different offered loads.

load. This effect appears to be associated with the stations with higher offered load making better use than the voice station of the transmission opportunities won by each station: the low-rate voice call may have no packet to send when a transmission opportunity is won, in which case the station needs to begin a new countdown to win a further transmission opportunity when a packet does arrive.

The results in Figure 7 are for a single hop. Evidently, if this effect is present over multiple hops it can have a very substantial cumulative impact on the throughput achieved by the voice call. Figure 8(a) shows how the voice throughput falls as the number of hops is increased and as the offered load on competing stations is varied. These results are for a simple linear topology with 5 stations competing with the voice call as it is relayed at each hop; the traffic injected by these stations is not relayed, only the voice call is forwarded. We can see that the throughput achieved by the voice call decreases as the number of hops increases. The level of decrease is strongly dependent on the offered load at the competing stations. When the load is less than or similar to that of the voice call (64Kbs) the decrease in the throughput of the voice call remains small, even after 10 hops. However, at slightly higher offered loads the unfairness noted above leads to a rapid decrease in throughput. For comparison, Figure 8(b) shows the throughput achieved by a competing station.

In one sense this is not a new observation: many authors have noted a decrease in end-to-end throughput as the hop count is increased. On the other hand, the insight provided by our model as to the detailed relative-load based nature of this effect and its cumulative impact does seem to be new. While a number of changes have recently been proposed to the 802.11 MAC to support expedited forwarding across multiple hops (e.g. see [10], [11] and references therein), the insight provided by the analysis here suggests that the potential may exist to
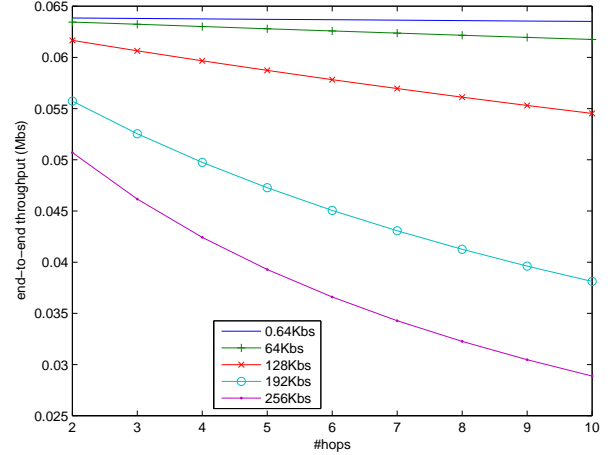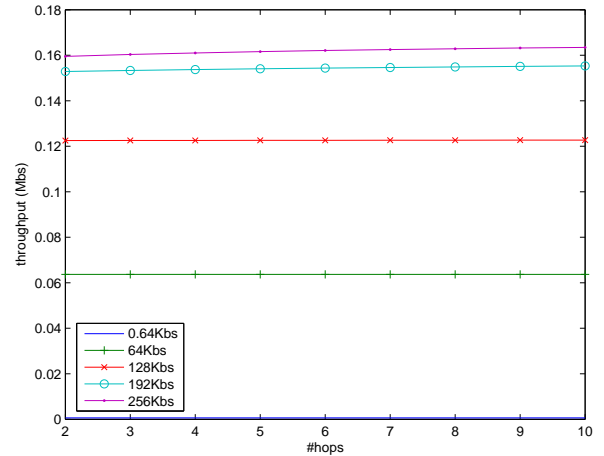
mitigate this problem using the flexibility provided by the new 802.11e standard. While we leave further consideration of this issue as future work, we illustrate the potential for improvement in Figure 9. This figure is for an identical scenario as Figure 8, but with the relay traffic prioritised by reducing the minimum contention window $CWmin$ used. It can be seen that the voice call throughput is now largely insensitive to the offered load at the contending stations at each hop and that the decrease in throughput, even after 10 hops, remains small (less than 10%).
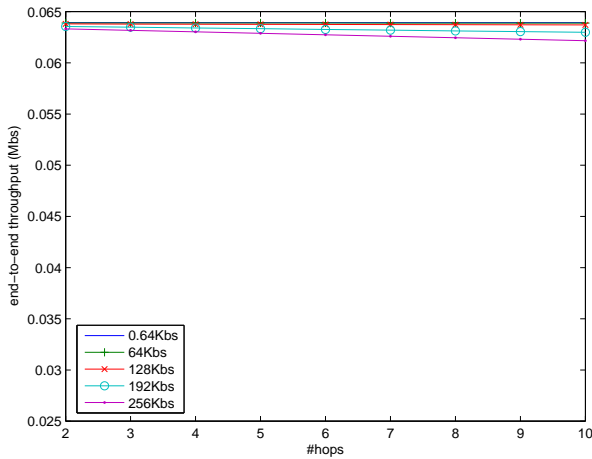
Fig. 9. Scenario as in Figure 8, but with relay traffic prioritised using a small value of $CWmin$.

VI. CONCLUSIONS

In this paper we introduce a tractable analytic model of throughput performance for general 802.11 multi-hop networks. We note that consideration of finite load is essential in a mesh network context as (i) relaying of traffic generally carries a contention and/or packet loss overhead at each hop; thus even if the stations at the first stage in a relay are saturated, stations at subsequent stages generally will not be. Modelling the scaling behavior of throughout with number of relay stages therefore makes consideration of finite loads essential. (ii) Realistic traffic such as voice and web is low-rate and on/off; thus network performance with such traffic cannot be accurately modeled without consideration of finite loads. We use this model to explore a number of fairness issues in 802.11 multi-hop networks that have a significant impact on performance and capacity for realistic traffic.

REFERENCES

[1] K. Duffy, D. Malone, and D. Leith, "Modeling the 802.11 Distributed Coordination Function in non-saturated conditions," *IEEE Communications Letters*, vol. 9, no. 8, pp. 715–717, 2005.

[2] D. Malone, K. Duffy, and D. Leith, "Modelling the 802.11 DCF under heterogenous finite load," in *Proc. RAWNET*, 2005.

[3] A. Markopoulou, F. Tobagi, and M. Karam, "Assessing the quality of voice communications over internet backbones," *IEEE Transactions on Networking*, vol. 11, no. 5, pp. 747–760, October 2003.

[4] N. Hegde, A. Proutiere, and J. Roberts, "Evaluating the voice capacity of 802.11 wlan under distributed control," in *Proc. IEEE LANMAN*, September 2005.

[5] P. Clifford, K. Duffy, D. Leith, and D. Malone, "On improving voice capaicty in 802.11 infrastructure networks," in *Proc. IEEE WirelessCom*, June 2005.

[6] S. Biswas and R. Morris, "Opportunistic routing in multi-hop wireless networks," in *Proc. IEEE/ACM SIGCOMM*, 2005.

[7] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. ACM MobiCom*, 2004.

[8] A. Raniwala and T. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. IEEE INFOCOM*, 2005.

[9] C. Barrett and D. Engelhart, "Analysing the short-term fairness of IEEE 802.11 in wireless multi-hop radio networks," in *Proc. IEEE MASCOTS*, 2002.

[10] D.Raguin, M.Kubisch, H.Karl, and A.Wolisz, "Queue-driven cut-through medium access in wireless ad hoc networks," in *Proc. IEEE WCNC*, 2004.

[11] S. Narayanan, P. Liu, and S. Panwar, "On the advantages of multi-hop extensions to the IEEE 802.11 infrastructure model," in *Proc. IEEE WCNC*, 2005.

[12] S. Lee, S. Banerjee, and B. Bhattacharjee, "The case for a multi-hop wireless local area network," in *Proc. IEEE INFOCOM*, 2004.

[13] X. Wang and K. Kar, "Cross-layer rate optimization in multi-hop aloha networks," in *2005 IEEE International Conference on Communications*, vol. 5, May 2005, pp. 2942–2946.

[14] Y. Sun, X. Gao, E. Belding-Royer, and J. Kempf, "Model-based resource prediction for multi-hop wireless networks," in *Proc. IEEE MASS*, 2004.

[15] Y. Gao, D. Chiu, and J. Lui, "The fundamental role of hop distance in IEEE 802.11 multi-hop ad hoc networks," in *Proc. ICNP*, 2005.

[16] Y. Wang and J. Garcia-Luna-Aceves, "Collision avoidance in multi-hop ad hoc networks," in *Proc. IEEE MASCOTS*, 2002.

[17] Y. Barowski, S. Biaz, and P. Agrawal, "Towards the performance analysis of IEEE 802.11 in multi-hop ad-hoc networks," in *Proc. IEEE WCNC*, 2005.

[18] Y. Barowski and S. Biaz, "The performance analysis of IEEE 802.11 under unsaturated traffic conditions," Auburn University, Tech. Rep. CSSE0-4-09, 2004.

[19] D. Leith and P. Clifford, "A self-managed distributed channel selection algorithm for WLANs," in *Proc IEEE RAWNET 2006*, April 2006.

[20] R. Battiti and B. Li, "Supporting service differentiation with enhancements of the IEEE 802.11 MAC protocol: models and analysis," University of Trento, Tech. Rep. DIT-03-024, May 2003.