Forensic Analysis of Packet Losses in Wireless Networks

Jianxia Ning*, Shailendra Singh*, Konstantinos Pelechrinis§, Bin Liu‡,
Srikanth V. Krishnamurthy*, and Ramesh Govindan‡

†University of California, Riverside: {jning, singhs, krish}@cs.ucr.edu, *University of Pittsburgh: kpele@pitt.edu, ‡University of Southern California: {binliu, ramesh}@usc.edu

Abstract—**Due to the lossy nature of wireless links, it is difficult to determine if packet losses are due to wireless-induced effects or from malicious discarding. Many prior efforts on detecting malicious packet drops rely on evidence collected via passive monitoring by neighbor nodes; however, they do not analyze the cause of packet losses. In this paper, we ask: (a) Given certain macroscopic parameters of the network (like traffic intensity and node density) what is the likelihood that evidence exists with respect to a transmission? and, (b) How can these parameters be used to perform a forensic analysis of the reason for the losses? Towards answering the above questions, we first build an analytical framework that computes the likelihood that evidence (we call this transmission evidence or TE for short) exists with respect to transmissions, in terms of a set of network parameters. We validate our analytical framework via both simulations as well as real-world experiments on two different wireless testbeds. The analytical framework is then used as a basis for a protocol within a forensic analyzer to assess the cause of packet losses and determine the likelihood of forwarding misbehaviors. Through simulations, we find that our assessments are close to the ground truth in all examined cases, with an average deviation of 2.3% from the ground truth and a worst case deviation of 15.0%.**

I. Introduction

Wireless ad hoc and mesh networks find application in municipal networks, tactical deployments and disaster recovery missions. In such networks, packet forwarding along a path is an inherent functional requirement. There have been studies on packet dropping attacks, wherein malicious routers that are required to forward packets do not do so (e.g. [1]). Unfortunately, due to the lossy nature of wireless links it is not easy to determine whether packet losses are due to natural wireless induced effects (channel impairments or interference) or due to such malicious drops.

Forensic systems typically collect evidentiary data towards detecting such packet dropping attacks (e.g. [2], [3]); however, they do not make any analysis to distinguish between wireless induced losses and malicious drops. Nodes that are part of the network themselves may act as witnesses and monitor transmissions [4]; this is an attractive option when networks are rapidly deployed and dedicated monitoring nodes are unavailable. Depending on the deployment, witnesses may not have evidence (e.g., due to very few witnesses or because of high levels of interference) relating to certain transmissions.

In this paper, our primary objective is to perform a forensic analysis on the cause of packet losses based on some macroscopic network parameters (such as traffic intensity and network density) in multi-hop wireless networks. Specifically, we seek to answer the questions: **(a)** Given a set of macroscopic network parameters, what is the likelihood that evidence exists relating to transmissions? and, **(b)** How can one perform a forensic assessment to determine if packet losses on links are due to natural effects in a wireless network or due to malicious discarding, based on these macroscopic network parameters?

978-1-4673-2447-2/12/$31.00 © 2012 IEEE

Towards answering the above questions, we construct an analytical framework that takes as input, macroscopic measurements or configurations of network properties (as alluded to above) and provides as output the probability that evidence exists relating to transmissions. We call this evidence, "transmission evidence" or TE for short. The analytical framework forms the basis for a protocol used within a forensic analyzer to assess the most likely cause of packet losses on links.

In particular, our contributions in this paper are as follows:

(i) Computing the likelihood of TE availability: We construct an analytical framework for computing the likelihood of TE availability. We capture the factors that affect TE availability on both individual links and on an end-to-end path. We find that the availability depends on network parameters such as packet size, bit-rates, traffic load and node density. We make several interesting observations on the trends in TE availability when tuning these parameters.

(ii) Validating the analytical models via simulations and real experiments: We perform extensive simulations to validate our analytical framework. We also perform experiments on (a) a 802.11 testbed and, (b) a testbed with five WARP boards [5] towards our validation. We find that our analytical framework can adequately capture the likelihood of TE availability in real networks.

(iii) Forensic analysis of packet losses: Our analytical framework facilitates the estimation of the likelihood of either a transmitter and/or a receiver discarding packets, given the conditions in the network. The framework is used as the basis for a protocol within a forensic analyzer. It takes as input (a) the network parameters and (b) monitoring logs for the considered link; it then yields the likelihood that the transmitter or the receiver on the link has discarded packets. We perform extensive simulations and compare the assessment results with ground truth. We find that our analyzer facilitates assessments with high accuracy; in particular, they deviate from the ground truth by 2.3%, on average.

Scope of our work: Our analytical models yield a quick and effective way of capturing the TE availability in large sets of scenarios. The advantage of the approach is that only a coarse estimate of network parameters is used in order to make the assessments. In this paper we have considered relatively static, homogeneous (e.g. a single packet size is used) settings; even for this, the construction of the analytical framework that forms the basis of our forensic analyzer, is non-trivial. More importantly, the models capture the trends in TE availability in practice as validated by both our simulations and experiments on real systems. A consideration of more complex settings is left to future work.

The output of our forensic analyzer provides coarse-grained assessments on forwarding misbehaviors. Because of the generality of the analytical framework applied therein, the assessments of the cause of packet losses on specific links inevitably deviate from the ground truth. However, our evaluations show that in all cases we examine, the average deviation from the truth is about 2.3%, while the maximum deviation is 15.0%.

We wish to point out here that since malicious drops are always supplementary to wireless induced losses, it is impossible for an attacker to exactly mimic natural wireless effects. The likelihood of an attack being detected will directly depend on the aggressiveness of the attacker; the more the drops, the more the deviation from what is expected due to natural wireless effects and thus, the higher the chance of detection.

Organization: The paper is structured as follows. Section II discusses related work. In Section III, we provide a description of our analytical framework. We consider specific network parameters and apply these in our framework in Section IV. The applicability of our framework in a forensic analyzer is discussed in Section V. In Section VI, we present our performance evaluations. We conclude in Section VII.

II. Related Work

In this section we briefly discuss related literature on network forensics and analyzing packet losses in wireless networks.

There is prior work on wireless monitoring at the *mechanism and system design* level [1], [6], [3], [7], [8], [9], [10], [11]. Marti *et al.* design a *watchdog* scheme to identify malicious nodes which do not forward packets along a multi-hop path. McGrath *et al.* [6] design and implement FLUX; FLUX automates the collection of forensic data and identifies abnormal traffic and network weaknesses. Ramach *et al.* [3] design and implement DAMON, a distributed monitoring system for MANETs. In summary, almost all of the above approaches propose techniques for solving specific network problems that require evidentiary data. None of them study the impact of various network parameters on the collection of evidence as we do here. Moreover, only a few use the evidentiary data to detect packet dropping attacks [3], [8]; however, unlike in these efforts we try to determine the likely cause of packet losses using a macroscopic view of network parameters.

ETX [12] and ETT [13] are metrics that have been designed to estimate the packet delivery ratio on links; however, they are empirical and more importantly these metrics reflect the packet loss rate but do not give insights into the root-cause of packet losses.

Some prior efforts attempt to distinguish packet losses due to interference from those due to channel fading [14], [15], [16]. Reis *et al.* present models for the physical layer behaviors of static wireless networks, focusing on the successful packet reception and carrier sense with interference. Qiu *et al.* propose a general model that is able to capture collision-induced losses in multihop wireless networks, based a limited number of measurements. Wong *et al.* propose Robust Rate Adaptation Algorithm (RRAA), by which they try to differentiate between fading-related and collision losses. None of these efforts however, consider the possibility of malicious discarding of packets.

The work that is closest to ours is in [4]. It proposes a specific witness-based detection scheme to identify forwarding misbehaviors. The authors analytically show that their scheme has low false positive and false negative rates. However, they do not evaluate how various network parameters would affect the evidence availability. To our best knowledge, we are the first to propose analytical models and experimental validation for this purpose.

III. Our Analytical framework

In this section, we develop our analytical framework to compute the likelihood of TE availability. At this time, we assume that neither the transmitter nor the receiver discards packets maliciously. We defer a discussion of how our framework can be applied in a forensic analyzer to identify such possibilities, to Section V.

Evidence maintenance: In a multi-hop static wireless network, nodes maintain evidence relating to transmissions as follows: **(a)** A sender (or transmitter) keeps the signed ACK it receives for each packet it sends. **(b)** A receiver creates an entry locally for each unique packet received and digitally verified. **(c)** A monitoring (witness) node creates an entry locally for each packet that it overhears and verifies. We assume that storage is not a limiting factor in evidence collection; one can envision nodes sending coarse-grained information relating to collected evidence periodically, to a central forensic controller. The overhead due to digital signatures has been previously studied (e.g., [17]) and hence is not considered here. The signed ACK helps in assuring non-repudiation. The sender can validate that it sent the packet in question and the receiver cannot deny receiving the packet. Without loss of generality, we assume that an ACK includes sender and receiver IDs and thus, an overheard ACK is of evidentiary value.

We expect that evidence is only sent infrequently to the controller since our objective is to investigate long-term effects. It can be easily piggybacked onto other control information (e.g. routing updates) and thus, we expect that the overhead is likely to be small. Our focus in this paper is more on the forensic analysis itself and not on the evidence collection process; thus, we do not perform an analysis of the overhead consumed due to evidence gathering.

Hop-level TE (HTE): The availability of hop-level TE reflects the likelihood that evidence exists relating to transmissions on a link. Commonly used notations are enlisted in Table I.

As stated above, for a transmission between v_i and v_j, there are three sources of evidence.

N	Total number of nodes
v_i	Transmitter
v_j	Receiver
d_{v_i,v_j}	Distance between v_i and v_j
P_{v_i,v_j}	Received power at v_j from v_i
h_{v_i,v_j}	Channel attenuation between v_j and v_i
η	Expected value of $\lvert h_{v_i,v_j}\rvert^2$
P_t	Transmission power
P_n	Noise power
α	Path loss exponent
z	Number of interferers
Z	Set of interferers
λ	Expected traffic sent per node in unit time
Λ	Expected interference level perceived by a node projected from another node in unit time
r	Transmission bit-rate
γ	SINR threshold
l_D	data packet length
l_A	ACK packet length

TABLE I: Notations

Source 1: v_i ***has the signed ACK from*** v_j ***for packet*** $(src, dest, pkt\ SQN)$. This requires that: **(a)** v_i's data packet is successfully received by v_j and, **(b)** v_j's ACK packet is successfully received by v_i.

$Pr\left(succ \mid r, l_D\right)$ denotes the probability of a successful data transmission with rate r and packet length l_D. Similarly, the probability of a successful ACK transmission is $Pr\left(succ \mid r_0, l_A\right)$, assuming that an ACK is sent at the base rate r_0 and has a length l_A. The probability that the first source of evidence is available is

$$Pr_{src1} = Pr\left(succ \mid r, l_D\right) \cdot Pr\left(succ \mid r_0, l_A\right). \tag{1}$$

Source 2: v_j ***has a stored entry*** $|v_i|src|dest|pkt\ SQN|$ $timestamp|$. This source of TE requires a successful transmission from v_i to v_j, the probability of which is

$$Pr_{src2} = Pr\left(succ \mid r, l_D\right). \tag{2}$$

Source 3: At least one witness has a stored entry $|v_i|src|$ $dest|pkt\ SQN|timestamp|$. This requires at least a node other than v_i or v_j to overhear the data transmission from v_i or the ACK transmission from v_j.

Let Pr_{src3_D} and Pr_{src3_A} denote the probabilities that at least one witness overhears the data and ACK, respectively. Note that due to the half-duplex property of typical radio devices, it is assumed that a node cannot be an interferer and a witness for the same transmission[1]. Therefore, when there are totally N nodes in the network and z interferers, the number of witnesses cannot exceed $N - z - 2$. Thus we have:

$$\begin{aligned} Pr_{src3\_D} = \sum_{z=0}^{N-2} & Pr\left(\mathrm{Z}\ \ int \mid r, l_D\right) \cdot \\ & \left(1 - \left(1 - Pr\left(succ \mid r, l_D, z\right)\right)^{N-z-2}\right), \end{aligned} \tag{3}$$

in which $\left(1 - \left(1 - Pr\left(succ \mid r, l_D, z\right)\right)^{N-z-2}\right)$ is the probability that given z interferers, at least one witness overhears the data transmission. Considering all possible values that z can take, we get the marginal probability that at least one witness overhears a given transmission.

In a similar way, we compute Pr_{src3_A} as follows[2]

$$\begin{aligned} Pr_{src3\_A} = Pr\left(succ \mid r, l_D\right) \cdot & \left(\sum_{z=0}^{N-2} Pr\left(\mathrm{Z}\ \ int \mid r_0, l_A\right) \cdot \right. \\ & \left. \left(1 - \left(1 - Pr\left(succ \mid r_0, l_A, z\right)\right)^{N-z-2}\right)\right). \end{aligned} \tag{4}$$

The successful overhearing of data and ACK transmissions by any given witness are assumed to be independent. In reality, there may be correlations due to interference effects at the overhearing node. However, the assumption (which we make for tractability) is shown to be reasonable by our simulations/experiments. With this:

$$Pr_{src3} = Pr_{src3\_D} + \left(1 - Pr_{src3\_D}\right) \cdot Pr_{src3\_A}. \tag{5}$$

Hop-level TE availability: The probability that at least one source of TE is available for a transmission, under the assumption of evidence independence[3] is

$$\begin{aligned} Pr_{HTE} &= 1 - \prod_{i=1}^{3} Pr(source\ i\ is\ unavailable) \\ &= 1 - (1 - Pr_{src1}) \cdot (1 - Pr_{src2}) \cdot (1 - Pr_{src3}). \end{aligned} \tag{6}$$

Accounting for retransmissions: Next we consider a limit of n_r retransmissions for the same data packet. The success of each transmission is independent from that of another (assuming that these are staggered in time, this is a reasonable assumption since the temporal network conditions are likely to change). A successful exchange of data and ACK packets results in the termination of retransmission attempts. This probability of successful exchange, denoted by Pr_{succ_ex} is

$$Pr_{succ\_ex} = Pr\left(succ \mid r, l_D\right) \cdot Pr\left(succ \mid r_0, l_A\right), \tag{7}$$

The probability that there are i retransmissions ($i + 1$ transmission attempts) is denoted as $Pr(rtx)$ and is given by:

$$Pr(rtx = i) = \begin{cases} Pr_{succ\_ex} \cdot (1 - Pr_{succ\_ex})^i & 0 \le i \le n_r - 1 \\ 1 - \sum_{j=0}^{n_r-1} Pr(rtx = j) & i = n_r \end{cases}.$$

Hence, the TE availability probability with a retransmission limit of n_r is:

$$Pr_{HTE}[n_r] = \sum_{i=0}^{n_r} Pr(rtx = i) \cdot \left(1 - (1 - Pr_{HTE})^{i+1}\right). \tag{8}$$

Path-level TE (PTE): Next we look at the path-level TE, i.e., the evidence relating to all transmissions on an end-to-end path. The TE availability on each hop along the path is assumed to be independent of that on the other hops. Again, in reality the TE availability across hops may be correlated but we make this assumption for tractability; our simulations and experiments (where there is correlation) verify that this assumption is indeed acceptable. The PTE requires the HTE

[1] Without loss of generality we assume that the monitoring devices or witnesses are active nodes in the network. It is easy to modify the analysis if evidence is collected only by passive monitoring nodes.

[2] For v_j to transmit an ACK, it must have successfully received the corresponding data packet.

[3] Dependencies between sources of evidence are discussed in Section V.

on all the hops of the path. The PTE for a H-hop path, denoted by $Pr_{PTE}[H]$, is given by:

$$Pr_{PTE}[H] = \prod_{h=1}^{H} Pr_{HTE}[at\ h^{th}\ hop]. \quad (9)$$

Bit-rate selection: Different bit-rates used on different hops will cause the TE availability on each hop to differ. The bit rate used on a link depends not only on the physical conditions (e.g., the distance between the communicating pair, the temporal fluctuations due to fading) but also on the rate adaptation algorithm in use. Given these, it is difficult to come up with a distribution for the bit rates used by nodes in a network. For simplicity, we assume that a bit rate is selected randomly from among the set of available rates. Note however, that our analysis can easily incorporate other distributions characterizing the usage of different bit rates. The probability of PTE availability is computed by considering all possible combinations of rates, on each hop of the path.

With (9), we see that the same parameters that affect HTE affect PTE. In addition, the hop count H impacts PTE; generally, as one may expect the longer the path, the lower PTE.

Note that the PTE as defined here is strict in the sense that it requires HTE on all hops. The TE of the transmission on hop h, can imply the success of transmissions on the previous $h-1$ hops, even though the HTE may not be available for all such hops. We will consider this sort of implicit PTE in future work.

IV. Explicitly Computing The Likelihood of TE availability

Now that we have computed the high level formulation of the likelihood of TE availability in Section III, we need compute the probabilities of success in (1), (2) and (5). However, in order to do so, we need to provide specific characteristics of the network. We proceed to do so in this section using commonly used models for representing the channel, the node density and the generated traffic; the models seem to characterize practical settings with good accuracy as seen in our real experiments later. Note here that, other models can be easily incorporated into our generic analytical framework.

The channel model: The received signal strength from node v_i, at node v_j is:

$$P_{v_i,v_j} = \frac{P_t \cdot |h_{v_i,v_j}|^2}{d^{\alpha}_{v_i,v_j}}, \quad (10)$$

where, P_t is the transmission power. h_{v_i,v_j} is the attenuation due to fading between the communicating pair. As typical, we assume that h_{v_i,v_j} is a Rayleigh distributed random variable [18]; thus, $|h_{v_i,v_j}|^2$ is exponentially distributed. d_{v_i,v_j} is the distance between v_i and v_j. α is the path loss exponent.

The collision model: There are several models used to capture collisions in the literature [19]. We use the *SINR (Signal-to-Interference-and-Noise) physical model*, where node v_j successfully receives the transmission from node v_i *iff*:

$$\frac{P_{v_i,v_j}}{P_n + \sum_{k\in\{1,...,N\}\backslash\{i,j\}} P_{v_k,v_j}} > \gamma, \quad (11)$$

where, P_{v_i,v_j} is the received power from v_i to v_j and is computed using (10). P_n is noise power. v_k is one of the interfering nodes. $\sum_{k\in\{1,...,N\}\backslash\{i,j\}} P_{v_k,v_j}$ is the accumulative interference power perceived by v_j. γ is the SINR threshold which varies with transmission bit-rate.

Use of multiple bit rates: The data packets are sent at a chosen transmission bit-rate from a set of available rates. For each rate there is a corresponding SINR threshold.

Media access control (MAC): To remove protocol dependencies, we do not assume a specific MAC scheme. Instead, we use a parameter to characterize the interference that nodes perceive, which in turn reflects the interference resolving ability of the MAC in use. This simplified representation avoids modeling the operations of specific MACs. As demonstrated later in Section VI, this model can be used to characterize multiple commonly used MAC protocols.

Node distribution: The network consists of N uniformly distributed static nodes (v_i, $i \in \{1, ..., N\}$).

Traffic pattern: Nodes send Poisson traffic, including their own packets and those to be simply forwarded.

Computing TE availability: Towards computing (6), we start by considering a transmission of a data packet from v_i to v_j. Given the distance between them d_{v_i,v_j}, the transmission bit-rate in use r, packet length l_D, and the number of interferers z, the probability of the transmission succeeding (denoted as $Pr\left(succ \mid r, l_D, d_{v_i,v_j}, z\right)$) is:

$$Pr\left(succ \mid r, l_D, d_{v_i,v_j}, z\right) = Pr\left(\frac{P_{v_i,v_j}}{P_n + \sum_{k\in Z} P_{v_k,v_j}} > \gamma\right), \quad (12)$$

where Z is the set of interferers. $Z \subset \{1, ..., N\}\backslash\{i, j\}$ and $|Z| = z$. The value of γ here, corresponds to the rate r in use.

With respect to the right hand side (RHS) of (12) there are two cases: (i) in the absence of interference (when $z = 0$) and, (ii) with the presence of interference (when $1 \leq z \leq N-2$). Detailed derivations of (12) for these two cases, are presented in an appendix.

Next, we remove the conditioning on the number of interferers z from $Pr\left(succ \mid r, l_D, d_{v_i,v_j}, z\right)$. Λ is a parameter that captures the expected interference at a given node from a neighbor node, per unit time. In reality this is dependent on both the traffic intensity and the MAC protocol in use. However, we try to capture the interference experienced at a node, simply with this parameter. If interference is managed (e.g. with TDMA or CSMA/CA), Λ is likely to be low. If the interference is unmanaged (as with say Aloha) and is high, Λ will be high. If we assume asynchronous transmissions and fixed sized packets, it is easy to see that a packet is interfered with, if another node initiates a transmission within the packet transmission time (say τ), or for a duration of τ prior to the beginning of the intended transmission (similar to the analysis of the Aloha medium access scheme in [20]). Thus, if the traffic load of a node is λ, the projected interference load can be characterized by 2λ. Hence in this specific case, $\Lambda = 2\lambda$. When the traffic load is Poisson, the probability that a transmission does not *overlap* with the intended transmission between v_i

and v_j is:

$$g(0, \Lambda \frac{l_D}{r}) = e^{-\Lambda \frac{l_D}{r}} \left(\Lambda \frac{l_D}{r} \right). \tag{13}$$

The probability that there are z $(0 \leq z \leq N-2)$ interference sources during the data packet transmission time $\frac{l_D}{r}$ is:

$$Pr\left(\text{z } \ int \mid r, l_D\right) = \left(1 - g(0, \Lambda \frac{l_D}{r})\right)^z \cdot \{g(0, \Lambda \frac{l_D}{r})\}^{(N-2-z)}. \tag{14}$$

The probability of a successful data transmission given the bit-rate and the packet length in use, and the distance between the communicating pair, is:

$$\begin{aligned} & Pr\left(succ \mid r, l_D, d_{v_i,v_j}\right) \\ & = \sum_{z=0}^{N-2} Pr\left(\text{z } \ int \mid r, l_D\right) \cdot Pr\left(succ \mid r, l_D, d_{v_i,v_j}, z\right). \end{aligned} \tag{15}$$

Next we remove the conditioning on d_{v_i,v_j} from $Pr\left(succ \mid r, l_D, d_{v_i,v_j}\right)$. As discussed, if one were to assume a uniform node deployment distribution, the PDF of d_{v_i,v_j} is $\frac{2d}{R^2}$. Thus:

$$Pr\left(succ \mid r, l_D\right) = \int_0^R Pr\left(succ \mid r, l_D, d_{v_i,v_j}\right) \frac{2d}{R^2} dd. \tag{16}$$

We emphasize that (14) and (16) can be easily modified to incorporate other distributions of interference levels and node deployments.

$Pr\left(succ \mid r, l_D, z\right)$ in (3), is obtained by removing the conditioning on d_{v_i,v_j} from $Pr\left(succ \mid r, \ l_D, d_{v_i,v_j}, z\right)$ (similar to that in (16)). Together with $Pr\left(\text{z } \ int \mid r, l_D\right)$ and $Pr\left(succ \mid r, \ l_D\right)$, we get Pr_{src1}, Pr_{src2} and Pr_{src3}, and finally the probability that hop-level TE is available (Pr_{HTE}) in (6).

V. Our Forensic Analyzer

Our analytical framework is used as the basis for a protocol within a forensic analyzer. Using the framework the analyzer computes offline, the probabilities of packet losses and TE availability under different conditions, in a benign setting on a link, based on a set of network parameters. It then compares these computed values with what is observed during network operations to estimate the likelihood of a transmitter or receiver discarding packets and lying about the same. As discussed earlier, the packet losses due to malicious dropping will always be in addition to what is experienced in benign settings due to wireless effects. The more aggressive an attacker, the more will be the deviation between what is observed and the expected number of packet losses in benign settings.

In this section, we describe our forensic analyzer in detail.

Performing the forensic analysis: As illustrated in Fig. 1, the forensic analyzer takes as inputs 1) the estimated probabilities from our analytical framework and 2) the evidence collected by nodes at runtime and the packet delivery ratios (PDRs) reported by the receivers. It outputs the assessment results on possible forwarding misbehaviors as discussed below.

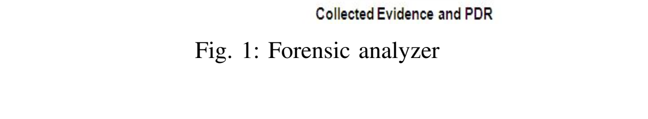

Measured or Configured Network Parameters
Analytical framework
Estimated TE and PDR
Forensic analyzer
Assessment
Collected Evidence and PDR

Fig. 1: Forensic analyzer

Forwarding misbehaviors: Nodes on an end-to-end path in a multi-hop wireless network may indulge in forwarding misbehaviors. A ***lying transmitter*** may claim to have attempted to forward packets, but may not have done so. Evidence for the transmissions that did not occur will not exist. A ***lying receiver*** may claim to have not received packets that were in fact received. If a receiver denies receiving packets, the only source of TE comes from any witness overhearing the data transmissions (available with probability Pr_{src3_D}). We wish to point out here that we do not differentiate between misbehaviors due to malicious activity and that from misconfigurations.

Threat model: In this work, we only consider forwarding misbehaviors as above. We assume that the network parameters are accurately gathered and nodes do not lie with regards to these parameters. We assume that keys cannot be compromised to create fake signatures. We also assume that there is no evidence manipulation i.e., none of the nodes create fake evidence or delete the genuine evidence. While a receiver discards packets as above, we assume it still follows the protocol in sending ACKs (only) for packets that it does not discard. Given these assumptions, the first source of evidence is conditional on the second source i.e., an ACK is possible only if the receiver says that it received the data packet successfully. Overheard ACKs, as part of the third source of evidence, are also dependent on the event that the receiver successfully receives the data packets. In other words, the first source of evidence and evidence with overheard ACKs will be available *iff* the second source of evidence is available (receiver has successfully recorded the data packet). With this, it is easy to see that (6) can be refined to $\{1 - (1 - Pr_{src2}) \cdot (1 - Pr_{src3\_D})\}$.

Analysis of misbehaviors: Suppose that $Pr[transmitter\ lying]$ is the likelihood of a transmitter lying about sending packets (which it does not send). Let $Pr\ [receiver\ lying]$ be the likelihood of a receiver lying of not receiving packets (when it discarded such received packets). With these forwarding misbehaviors, the likelihood of TE availability is:

$$\begin{aligned} & \tilde{Pr}_{HTE} = 0 \cdot Pr[transmitter\ lying] + \\ & 1 \cdot (1 - Pr[transmitter\ lying]) \cdot Pr_{succ} \cdot (1 - Pr[receiver\ lying]) + \\ & Pr_{src3\_D} \cdot (1 - Pr[transmitter\ lying]) \cdot Pr_{succ} \cdot Pr[receiver\ lying] + \\ & Pr_{src3\_D} \cdot (1 - Pr[transmitter\ lying]) \cdot (1 - Pr_{succ}), \end{aligned} \tag{17}$$

where, Pr_{succ} is simply a shortened notation for $Pr\left(succ \mid r, l_D\right)$. The terms in the summation on the RHS of (17) correspond to the TE availability over all possible combinations of the transmitter and the receiver lying as detailed in Table II. As discussed, if the transmitter is lying, no TE is available. If the receiver is lying, witnesses may or may not have evidence to the transmission.

Case	TE availability probability
Transmitter lying	0
Transmitter not lying, receiver receiving the packet and not lying	1
Transmitter not lying, receiver receiving the packet and lying	Pr_{src3_D}
Transmitter not lying, receiver not receiving the packet	Pr_{src3_D}

TABLE II: TE availability under all possible cases

If $Pr[transmitter\ lying]$ and $Pr[receiver\ lying]$ are set to 0,

(17) reduces to $\{1-(1-Pr_{src2})\cdot(1-Pr_{src3\_D})\}$, which is exactly the TE availability in benign settings.

If a transmitter or/and receiver indulges in forwarding misbehaviors, the PDR reported by the receiver is affected. Only those packets that are sent by the transmitter, successfully received and truthfully reported by the receiver are counted towards successful delivery. This PDR is expressed as:

$$PDR = \\ (1 - Pr[transmitter\ lying]) \cdot Pr_{succ} \cdot (1 - Pr[receiver\ lying]). \tag{18}$$

Solving (17) and (18) yields $Pr[transmitter\ lying]$ and Pr $[receiver\ lying]$ as follows:

$$Pr[transmitter\ lying] = \\ 1 - \frac{\tilde{Pr}_{HTE} - PDR + PDR * Pr_{src3\_D}}{Pr_{src3\_D}}. \tag{19}$$

$$Pr[receiver\ lying] = \\ 1 - \frac{PDR \cdot Pr_{src3\_D}}{Pr_{succ} \cdot (\tilde{Pr}_{HTE} - PDR + PDR \cdot Pr_{src3\_D})}. \tag{20}$$

From (19) and (20), we see that there are four values essential towards computing the desired probabilities. First, one would need the measured actual TE availability and reported PDR from the network during operations. The probability $\tilde{Pr}_{HTE}$ is simply the ratio of the number of packets for which evidence is available to the total number of packets the transmitter claims to have sent. Pr_{src3_D} and Pr_{succ} are obtained from the analytical models. Using these, the desired probabilities for the setting are computed. Finally, the probability of packet losses due to either the transmitter lying or the receiver lying is $\{Pr[transmitter\ lying] + (1 - Pr[transmitter\ lying]) \cdot Pr[receiver\ lying]\}$; The complementary probability to this yields the likelihood of the losses being because of natural effects (channel induced or interference) in the wireless network.

Discussion: We wish to acknowledge here that the actual TE availability on specific hops, even without any forwarding misbehaviors, may vary from that predicted by our analytical framework. Our assessment may inevitably deviate from the ground truth. However, the approach provides a quick and coarse-grained estimation on the likelihood of forwarding misbehaviors. In Section VI, we find via simulations that our assessments do not deviate much from the ground truth. For further fidelity, the local traffic and topology in the proximity of a link of interest can be considered and the analysis modified for that setting; however, note that this would increase the volume of information collected towards performing the forensic analysis (since microscopic information from local neighborhoods are needed).

VI. Evaluations

In this section, we first validate our analytical framework (in benign settings) with both simulations, and experiments on two different testbeds. We also examine the trends in TE availability by varying different network parameters. These provide an understanding of the likelihood of the existence of TE in various settings. Finally, we conduct a forensic analysis of packet losses to assess the likelihood of forwarding misbehaviors via simulations.

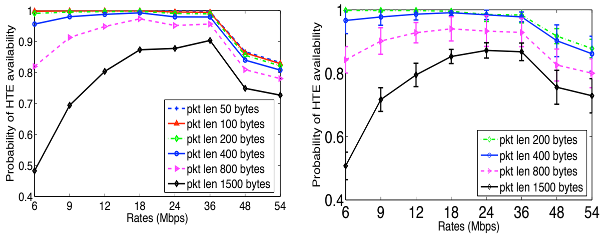
The default parameter settings (unless specified otherwise) are listed in Table III. Without loss of generality, the values for the rates and SINR thresholds are adopted from 802.11a.

N	10
P_t	3.16E-2 $watts$
P_n	3.16E-10 $watts$
λ	20 pkt/sec
R	100 m
η	1.0
α	2.0
Data packet length	50/100/200/400/800/1500 bytes
ACK length	20 bytes
Rates and SINR thresholds	See Table IV for the SINR thresholds and rates.

TABLE III: Default parameter settings

Rate	6	9	12	18	24	36	48	54
SINR	6.02	7.78	9.03	10.79	17.04	18.8	24.05	24.5

TABLE IV: 802.11a Rates (Mbps) and SINR thresholds (dB).

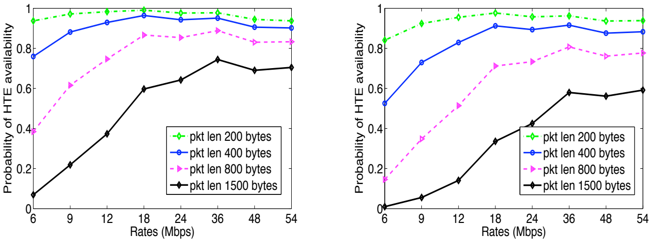


(a) Analytical (b) Simulation

Fig. 2: Hop-level TE availability probability

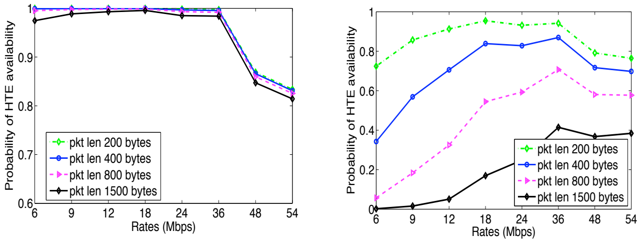
A. Model Validation via Simulations

Simulation setup: The simulations are performed using OPNET modeler version 14.5 [21]. In our simulations, we first consider a single hop wireless network. Here, N nodes are uniformly distributed in a circle with diameter $2R$. The considered receiver is positioned at the center, while the transmitter is randomly picked from among the other $N-1$ nodes. The transmissions experience both path loss and Rayleigh fading. Next, we consider a multi-hop network by spreading $5N$ nodes uniformly in a circle with diameter $2\sqrt{5}R$. The paths whose PTE is considered, are selected such that the nodes on the paths are near the center of the circle, instead of being at the network edge. We choose these configurations to eliminate edge effects while keeping the node density of the network fixed. Traffic are sent between randomly chosen source destination pairs. The traffic generated at a node is 20 pkts/sec. Shortest path routes are used. Nodes transmit packets at random instances in time to contend for channel access. At the end of a run, we combine the traces from all nodes and calculate the number of unique transmissions recorded. The fraction of this number over the total number of transmissions occurring during this run, is the TE availability probability. The data collected for each specific scenario is an average over 20 runs.



(a) Node number 20 (b) Node number 30

Fig. 3: Impact of node density



(a) Low Traffic Volume (b) High Traffic Volume

Fig. 4: Impact of traffic volume

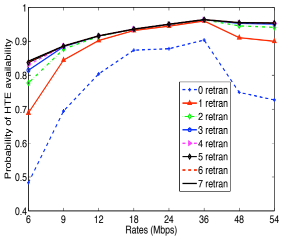


Fig. 5: Impact of retransmission limit (pkt len 1500 bytes)

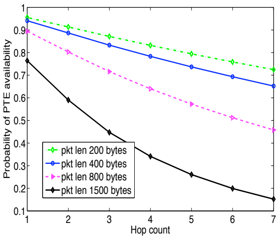


Fig. 6: PTE availability probability (analytical)

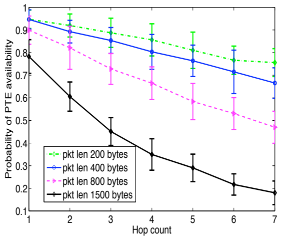


Fig. 7: PTE availability probability (simulation)

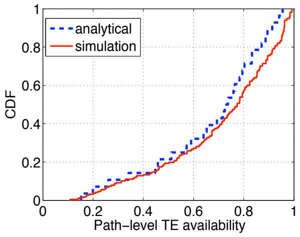


Fig. 8: CDF of PTE availability

Trends in hop-level TE availability: We first examine the trends in HTE availability when various parameters are tuned. Benign settings are considered.

Bit-rate and packet length: We first vary the transmission bit-rate and packet length. The other parameters are at default settings. Fig. 2 shows the trend in HTE availability from the analysis and simulations[4]. We see that for a fixed bit-rate, a smaller packet length leads to higher hop-level TE availability. This is because, with a smaller packet length, the air time is small; thus, the chances of a packet being corrupted due to interference either at the receiver or at the witnesses is lower. Furthermore, using lower bit rates results in almost perfect TE availability when the packet length is small. When the rate increases the TE availability decreases. With higher rates, packets are more susceptible to channel induced losses. This decreases the probabilities that the receiver and the overhearing witnesses successfully receive the transmissions. The above effects are more pronounced with larger packet lengths. In the case we consider, when the packet length grows beyond 800 bytes, even the lower rates do not guarantee high TE availability.

The results from simulations are shown in Fig. 2(b). We observe that the trends hold in terms of TE availability , thus validating the applicability of our assumptions.

Node density: We increase the number of nodes deployed from 10 to 20 and 30. As a consequence the node density increases. The analytical results are in Fig. 3 (simulation results are similar and not shown for purposes of clarity). The total interference levels imposed on a node is higher due to the higher node density. This hurts both packet reception and overhearing and eventually hurts TE. However, a high node density means that there are more nodes serving as potential witnesses. This helps TE collection. Figs. 3(a) and 3(b) indicate that (a) when low rates are used, the *first* factor seems dominant and, thus TE availability tends to decrease as N increases; (b) when high rates are used, the *second* factor seems dominant and TE availability increases with N.

Traffic volume: Now we adjust the traffic generated per node to be 5 times less (4 pkts/sec) and 5 times more (100 pkts/sec). As we see from Fig. 4(a), TE availability is fairly high and alike across all packet lengths with low traffic volume. It is because the main reason for packet failure is the effect of the channel and not the interference. In addition, with less traffic to send, nodes are more likely to be witnesses and collect TE. When traffic volume is high (Fig. 4(b)), TE availability drops drastically; high interference hurts TE collection and nodes have less time for overhearing.

Retransmission limit: We vary the retransmission limit from 0 (default) to 7. We notice that at low and moderate loads allowing more retransmissions increases TE availability as one might expect (Fig. 5). The TE availability drops when the retransmission load increases beyond a certain point. Note here that node density is a factor in determining when such a switch over would occur. Due to space constraints, we do not discuss more details here.

Trends in path-Level TE availability: We vary the packet length and hop count; other parameters remain at default settings. We look at paths with hop counts from 1 to 7. Recall that we assume a uniform rate selection at each hop. The results in terms of PTE availability probability generated with our analytical models and from simulations, are in Figs. 6 and 7 respectively. We observe that: (1) A shorter packet length yields a higher PTE availability. This trend is consistent with that in HTE. (2) The PTE decreases quite fast with increasing hop count. With the hop count being increased by 1, the PTE drops by 10%-20%. (3) The simulation results are similar to the analytical results.

Fig. 8 presents the CDFs of the PTE for all the cases that we examine. We observe in about 80% of the cases, the PTE is above 0.5; in about 30% of the cases, it exceeds or approaches

[4] The simulation results for packet lengths of 50 and 100 bytes are similar to that of length 200 bytes; they are not shown for clarity.

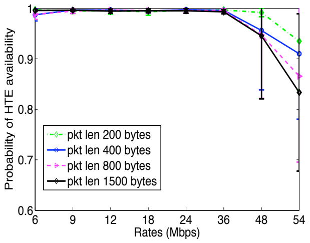


Fig. 9: Empirical HTE in 802.11.

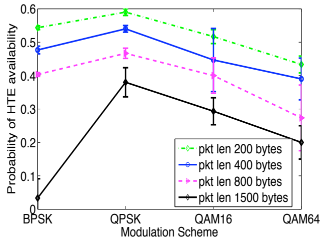


Fig. 10: Empirical HTE in Aloha.

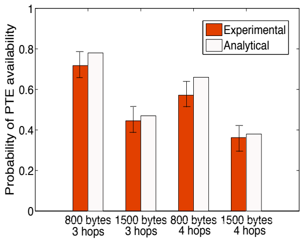


Fig. 11: Empirical PTE in 802.11.

0.8. This implies that, there is a good likelihood that the PTE is available in typical settings.

B. Model Validation via Experiments

We examine the TE availability in real networks: (a) in a 802.11 testbed with CSMA/CA and (b) in a testbed of WARP nodes with Aloha. Our objective is to show that our framework accurately characterizes the TE availability that one can expect in real networks.

Experiment setup: We conduct two sets of experiments, to examine the TE availability in *interference-managed* (802.11 CSMA/CA) and *-unmanaged* (Aloha) scenarios, respectively.

The first set of experiments is performed on a 42-node wireless testbed deployed on an entire floor of a campus building. The nodes are based on the Soekris net5501 hardware configuration, and run a Debian Linux distribution. We experiment with the 802.11a mode in order to avoid interference from co-located 802.11b campus WLANs. RTS/CTS is disabled. Each scenario involves different sets of about ten nodes. In each scenario, every node generates traffic to a randomly chosen neighbor, while running TCPdump to record all packets it receives or overhears. The TE availability probability is computed as described earlier with our simulations.

The second set experiments is on a 5-node WARP (Wireless open-Access Research Platform) [5] Radio testbed in a lab setting. Nodes access the channel using Aloha, which runs on top of the WARP OFDM implementation. Traffic is generated on the board itself. When a node is idle (not transmitting) it logs the data and ACK transmissions going on in its neighborhood and these logs are continuously sent to the central server which is connected to all the nodes using an Ethernet interface provided on each node. The central server is a PC which is basically used to send control messages to the nodes and to collect logs sent by the nodes. We evaluate the trends in TE using this testbed by varying packet lengths and modulation schemes. There is no option to change the forward error correction or FEC, code rate on our boards. The packet generation rate is 200 pkts/sec. We use 3 transmitters and 2 receivers for a period of 10 seconds.

Empirical hop-level TE: First, we report our experiments on the 802.11a testbed. We plot the TE availability probability for various packet lengths and with different rates in Fig. 9. We see that low rates offers high TE availability. At higher rates, the TE availability drops slightly ($\approx 10\%$), especially for larger packet sizes. A quick look at Fig. 4(a) shows that the results with our model with low traffic volume is similar to what is seen here. This is because, CSMA/CA manages the interference well to avoid interference in the vicinity of an active transmission. There is still interference due to hidden terminals, but the levels are low. Thus, by calibrating our model with a low Λ one can obtain trends that are likely to exist with MAC protocols that manage interference (as with 802.11 or TDMA).

Next, we consider unmanaged interference with Aloha. We use our WARP testbed here. Since, our testbed consists of only five nodes, we use a high packet generation rate in order to have a desired interference level.

Fig. 10 shows the TE trends with the four modulation schemes with Aloha. With our WARP boards, the FEC code rate is fixed and hence, there are not as many bit-rates to select from. The fairly limited setting of the testbed makes it difficult to directly compare the results with the heavy load scenario in Fig. 4(b), although the trends are similar. At low rates the TE availability is low. As we increase the rate it increases and finally drops due to channel induced effects with QAM 64 modulation. Note that there is no one to one mapping between the x-axes in Figs. 10 and 4(b). Thus, it is difficult to get an exact match. However, the model does indeed predict the trend of what could be expected in practice.

Empirical path-level TE: We use our 802.11 wireless testbed to validate the model for PTE. We create a number of 3 and 4-hop paths to measure the probability of end-to-end TE availability. Note that due to the small scale, such an experiment was not possible with our WARP hardware. We used static routing to create multi-hop paths to ensure that route flapping did not happen. Default rate adaptation is used by nodes. Each node along the path logs the traffic which is being transmitted in its vicinity, using which we calculate the PTE for each multi-hop path. Two packet sizes are considered. It is seen in Fig. 11 that the PTE availability decreases as the hop count and packet length increase. We see a good match with our analytical results generated with a low traffic volume.

Summary: To summarize, both our simulation and experimental results demonstrate that by appropriately calibrating our analytical framework with network parameters (packet size, bit rate in use, node density, interference level), one can get a good indication of the likelihood of TE availability in practice. This can not only aid forensic analysis (as discussed next), but also allow a network administrator to determine the efficacy of a monitoring system given specific network conditions.

C. Forensic Analysis Using TE

In our last set of evaluations, we aim to provide the assessments of forwarding misbehaviors. We randomly select up to one hundred links from our simulated network and on each of these links, ten thousand packets are scheduled to be

Ground	Assessment Results		
Truth (%)	avg dev (%)	min dev (%)	max dev (%)
transmitter 0	2.65	0.39	5.98
receiver 0	4.84	0.44	13.83
transmitter 10	2.05	0.00	1.05
receiver 0	4.20	0.00	11.77
transmitter 0	5.28	0.02	10.39
receiver 10	2.71	0.00	9.62
transmitter 10	1.80	0.18	6.08
receiver 10	1.84	0.38	5.88
transmitter 20	5.36	0.32	15.00
receiver 0	2.92	0.00	11.21
transmitter 0	1.36	0.00	8.49
receiver 20	2.32	0.35	9.24
transmitter 20	1.56	0.38	5.11
receiver 20	1.66	0.55	5.09
transmitter 40	3.87	0.29	10.89
receiver 0	2.59	0.00	9.09
transmitter 0	0.76	0.00	4.90
receiver 40	1.38	0.09	7.53
transmitter 40	0.99	0.26	3.11
receiver 40	1.01	0.16	3.05
transmitter 60	2.34	0.24	5.52
receiver 0	1.84	0.00	8.85
transmitter 0	0.23	0.00	1.42
receiver 60	1.89	0.50	5.20
transmitter 60	0.39	0.12	1.28
receiver 60	0.43	0.09	1.34

TABLE V: Assessments on transmitter and receiver lying

sent. We emulate forwarding misbehaviors at the transmitters and receivers, individually and jointly. A transmitter lying by $x\%$, implies that it does not transmit $x\%$ of the packets that it is supposed to send. A receiver lying by $x\%$ means that it claims to have received only $1-x\%$ of the packets that it in fact receives. We vary the fraction of lying (0%, 10%, 20%, 40% and 60%) at the transmitter and receiver, respectively. These preset values correspond to the ground truth.

We collect the actual TE and PDR for each transmission period. Having the measured values from the simulations and their estimated counterparts from the analytical models, we use our forensic analyzer described in Section V to assess the likelihoods of the transmitter and/or the receiver lying.

The assessment results are presented in Table V. Column one contains the ground truth, while columns two to four contain the average/minimum/maximum deviation of the assessments from the truth (expressed as percentages) across all considered links. The deviation is calculated as $|assessed\ value\ -\ truth|$ and is computed for both the transmitter and receiver. We see that, overall, our assessments are able to reflect the ground truth with good accuracy. However, due to the variance between the generic analytical models and the unique circumstances of each link, inevitably there is a deviation in the assessment on each specific link. The average deviation is 2.3% for all the cases that we examine, while the maximum value is 15.0%. These results demonstrate that our analytical framework can facilitate the assessment of the considered misbehaviors with good accuracy.

Note here that the deviation, as we define here, computes the "overestimate" or the "underestimate" of the misbehavior probability by the forensic analyzer in absolute terms. If this deviation is small, the analyzer has a reasonable estimate of the likelihood of misbehaviors. If needed, it can further gather fine grained information from the vicinity of the link in question to refine this probability estimate.

VII. Conclusions

In this paper, we seek to differentiate between wireless induced packet losses and malicious discarding in wireless networks. Towards facilitating such a forensic analysis, we develop an analytical framework that takes as input various macroscopic network parameters and yields as output, the likelihood of evidence availability. We validate our analytical framework via both extensive simulations and experiments on two different wireless testbeds that employ different MAC protocols. We then discuss the applicability of our analytical framework in a forensic analyzer to determine the likelihood of a transmitter or receiver discarding packets maliciously. We show via simulations that the analyzer is able to determine these likelihoods with high accuracy.

Acknowledgement: This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] S. Marti, T. J. Giuli, K. Lai, and M. Baker. Mitigating routing misbehavior in mobile ad hoc networks. In *ACM MOBICOM*, 2000.

[2] K. P. McGrath and J. Nelson. Monitoring & Forensic Analysis for Wireless Networks. In *Conf. on Internet Surveillance and Protection*, 2006.

[3] K. N. Ramach, E. M. Belding-royer, and K. C. Almeroth. Damon: A distributed architecture for monitoring multi-hop mobile networks. In *IEEE SECON*, 2004.

[4] S. Yang, S. Vasudevan, and J. Kurose. Witness-based detection of forwarding misbehaviors in wireless networks. In *UMass Computer Science Technical Report UM-CS-2009-001*, 2009.

[5] Wireless Open-Access Research Platform. http://warp.rice.edu/.

[6] K. P. McGrath and J. Nelson. Flux: A forensic time machine for wireless networks. In *IEEE INFOCOM 2006 Poster and Demo Session*, 2006.

[7] A. Adya, P. Bahl, R. Chandra, and L. Qiu. Architecture and techniques for diagnosing faults in ieee 802.11 infrastructure networks. In *ACM MOBICOM*, 2004.

[8] L. Qiu, P. Bahl, A. Rao, and L. Zhou. Troubleshooting wireless mesh networks. *ACM SIGCOMM Computer Communication Review*, 2006.

[9] J. Yeo, M. Youssef, and A. Agrawala. A framework for wireless lan monitoring and its applications. In *ACM workshop on Wireless security: WiSe*, 2004.

[10] Y.-C. Cheng, J. Bellardo, P. Benko, A. C. Snoeren, G. M. Voelker, and S. Savage. Jigsaw: Solving the puzzle of enterprise 802.11 analysis. In *SIGCOMM*, 2006.

[11] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Analyzing the mac-level behavior of wireless networks in the wild. In *SIGCOMM*, 2006.

[12] D. S. J. De Couto, D. Aguayo, J. Bicket, and R. Morris. A high-throughput path metric for multi-hop wireless routing. In *ACM MOBICOM*, 2004.

[13] R. Draves, J. Padhye, and B. Zill. Routing in multi-radio, multi-hop wireless mesh networks. In *ACM MOBICOM*, 2004.

[14] C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Measurement-based models of delivery and interference in static wireless networks. In *ACM SIGCOMM*, 2006.

[15] L. Qiu, Y. Zhang, F. Wang, M. Han, and R. Mahajan. A general model of wireless interference. In *ACM MOBICOM*, 2007.

[16] S. H. Y. Wong, H. Yang, S. Lu, and V. Bharghavan. Robust rate adaptation for 802.11 wireless networks. In *ACM MOBICOM*, 2006.

[17] S. S. Yau, Y. Yin, and H. G. An. An adaptive approach to optimizing tradeoff between service performance and security in service-based systems. In *International Journal of Web Services Research*, 2011.

[18] T. S. Rappaport. *Wireless Communications: Principles and Practice (2nd ed.).* Prentice Hall, 2001.

[19] C-K. Chau, M. Chen, and S. C. Liew. Capacity of large-scale csma wireless networks. In *ACM MOBICOM*, 2009.

[20] D. Bertsekas and R. Gallager. *Data Networks.* Prentice Hall, New Jersey, 1992.

[21] Opnet User's Documentation. http://www.opnet.com.

[22] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw–Hill, New York, 1991.

APPENDIX

The notation used here is carried over from Table I.

There are two cases when considering (12). In the absence of interference,

$$\begin{aligned} &Pr\left(succ \mid r, l_D, d_{v_i,v_j}, z=0\right) = Pr\left(\frac{P_{v_i,v_j}}{P_n} > \gamma\right) \\ &= Pr\left(|h_{v_i,v_j}|^2 > \frac{d^{\alpha}_{v_i,v_j} \cdot P_n \cdot \gamma}{P_t}\right). \end{aligned} \tag{21}$$

We denote $\frac{d^{\alpha}_{v_i,v_j} \cdot P_n \cdot \gamma}{P_t}$ by c. Recall that $|h_{v_i,v_j}|^2$ is an exponentially distributed r.v. with parameter η. Thus,

$$Pr\left(|h_{v_i,v_j}|^2 > c\right) = \int_c^{\infty} \eta e^{-\eta x} dx. \tag{22}$$

With interference, the success probability is computed as:

$$\begin{aligned} &Pr\left(succ \mid r, l_D, d_{v_i,v_j}, 1 \le z \le N-2\right) \\ &= Pr\left(\frac{P_{v_i,v_j}}{P_n + \sum\limits_{k \in Z} P_{v_k,v_j}} > \gamma\right) \\ &= Pr\left(\frac{P_t \cdot |h_{v_i,v_j}|^2}{d^{\alpha}_{v_i,v_j}} > \gamma \cdot \sum_{k \in Z} \frac{P_t \cdot |h_{v_k,v_j}|^2}{d^{\alpha}_{v_k,v_j}} + \gamma \cdot P_n\right) \end{aligned}$$

It is difficult to compute the above since the distances from v_j, to different interferers will be different. For tractability, we make a conservative approximation that all the interferers are at the same distance as that to the closest interferer to v_j[5] (denoted as $min\{d_{v_k,v_j}\}$). With this, we find a lower bound on the success probability (upper bound on failure probability) as follows:

$$\begin{aligned} &Pr\left(succ \mid r, l_D, d_{v_i,v_j}, 1 \le z \le N-2\right) \\ &> Pr\left(\frac{P_t \cdot |h_{v_i,v_j}|^2}{d^{\alpha}_{v_i,v_j}} > \gamma \cdot \sum_{k \in Z} \frac{P_t \cdot |h_{v_k,v_j}|^2}{min\{d_{v_k,v_j}\}^{\alpha}} + \gamma \cdot P_n\right) \\ &= Pr\left(P_t \cdot |h_{v_i,v_j}|^2 \cdot min\{d_{v_k,v_j}\}^{\alpha} > \right. \\ &\left. \gamma \cdot P_t \cdot \sum_{k \in Z} |h_{v_k,v_j}|^2 \cdot d^{\alpha}_{v_i,v_j} + \gamma \cdot min\{d_{v_k,v_j}\}^{\alpha} \cdot P_n \cdot d^{\alpha}_{v_i,v_j}\right). \end{aligned} \tag{23}$$

On the RHS of (23), there are three r.v.s $min\{d_{v_k,v_j}\}^{\alpha}$, $\sum_{k \in Z} |h_{v_k,v_j}|^2$ and $|h_{v_k,v_j}|^2$. Since $|h_{v_k,v_j}|^2$ is exponentially distributed with parameter η, the sum $\sum_{k \in Z} |h_{v_k,v_j}|^2$ follows an Erlang distribution with parameters η and z [22]. Next we derive the distribution of $min\{d_{v_k,v_j}\}^{\alpha}$ (d^{α}_{min} for short):

$$F_{d_{min}}(d) = Pr(d_{min} \le d) = 1 - Pr(d_{min} > d). \tag{24}$$

Recall that d_{min} is the minimum of the distances from the z interferers to v_j (denoted by $d_1, d_2, \cdots, d_z$). Thus,

$$\begin{aligned} 1 - Pr(d_{min} > d) &= 1 - Pr(min\{d_1, d_2, ..., d_z\} > d) \\ &= 1 - Pr(d_1 > d, d_2 > d, ..., d_z > d). \end{aligned} \tag{25}$$

Since $d_1, d_2, \cdots d_z$ are independent r.v.s with the same distribution, (25) can be written as:

$$\begin{aligned} &1 - Pr(d_1 > d) \cdot Pr(d_2 > d) \cdots Pr(d_z > d) = 1 - Pr(\hat{d} > d)^z \\ &= 1 - \left(1 - Pr(\hat{d} \le d)\right)^z = 1 - \left(1 - F_{\hat{d}}(d)\right)^z \end{aligned} \tag{26}$$

where $\hat{d}$ is the distance from an interferer to v_j. The probability distribution of $\hat{d}$ is simply the distribution of distance between node pairs in the network (since the interferer could be anywhere within the range of v_j). If one assumes a uniform deployment of nodes, the probability density function (PDF) that a node is d units away from another node is $\frac{2d}{R^2}$ where R is the maximum possible distance units between a pair.

The PDF of d_{min}, $f_{d_{min}}(d)$ is then given by:

$$\begin{aligned} f_{d_{min}}(d) &= \frac{d}{dd} F_{d_{min}}(d) = z \cdot (1 - F_{\hat{d}}(d))^{(z-1)} \cdot f_{\hat{d}}(d) \\ &= z \cdot \left(1 - \frac{d^2}{R^2}\right)^{z-1} \cdot \frac{2d}{R^2}. \end{aligned} \tag{27}$$

The PDF of d^{α}_{min} (denoted as $\tilde{d}$), a function of d_{min}, is expressed as [22]:

$$f_{\tilde{d}}(d) = \frac{1}{\alpha} \cdot d^{(\frac{1}{\alpha} - 1)} \cdot f_{d_{min}}(d^{\frac{1}{\alpha}}). \tag{28}$$

We now have the PDFs for the three r.v.s $min\{d_{v_k,v_j}\}^{\alpha}$, $\sum_{k \in Z} |h_{v_k,v_j}|^2$ and $|h_{v_i,v_j}|^2$. To compute (23), we need to further get the PDFs for $\{P_t \cdot |h_{v_i,v_j}|^2 \cdot min\{d_{v_k,v_j}\}^{\alpha}\}$ (denoted as a new r.v. V) and $\{\gamma \cdot P_t \cdot \sum\limits_{k \in Z} |h_{v_k,v_j}|^2 \cdot d^{\alpha}_{v_i,v_j} + \gamma \cdot min\{d_{v_k,v_j}\}^{\alpha} \cdot P_n \cdot d^{\alpha}_{v_i,v_j}\}$ (denoted as r.v. U).

We start by computing the PDF of $\{|h_{v_i,v_j}|^2 \cdot min\{d_{v_k,v_j}\}^{\alpha}\}$ (denoted as r.v. W). Note that $|h_{v_i,v_j}|^2$ (Y for short) and $min\{d_{v_k,v_j}\}^{\alpha}$ ($\tilde{d}$) are independent. Further, $\tilde{d}$ varies from 0 to $min(R^{\alpha}, w/y)$ where, w and y are variables representing the value assumed by r.v.s W and Y, respectively. Thus,

$$F_W(w) = \int_{\frac{w}{R^{\alpha}}}^{\infty} \int_0^{\frac{w}{y}} f_Y(y) f_{\tilde{d}}(d) dd\, dy + \int_0^{\frac{w}{R\alpha}} \int_0^{R^{\alpha}} f_Y(y) f_{\tilde{d}}(d) dd\, dy. \tag{29}$$

Differentiating (29) yields $f_W(w)$. Since $V = P_t \cdot W$:

$$f_V(v) = \frac{1}{P_t} \cdot f_W(\frac{v}{P_t}). \tag{30}$$

The derivation of the PDF of U is similar to that of V and is omitted due to space constraints.

With the new notation, the RHS of (23) is:

$$Pr(V > U) = \int_0^{\infty} \int_0^{v} f_V(v) f_U(u) du\, dv. \tag{31}$$

Having derived (22) and (23), we have the expression for $Pr\left(succ \mid r, \;\; l_D, d_{v_i,v_j}, z\right)$.

[5] Our simulations validate that this is reasonable.