Analysis and Performance of Topology Inference in Mobile Ad Hoc Networks

J. David Brown(✉), Mazda Salmanian, and Tricia J. Willink

Defence Research and Development Canada, Ottawa, Canada
{david.brown,mazda.salmanian,tricia.willink}@drdc-rddc.gc.ca

Abstract. This paper examines the performance of a strategy for mapping the topology of a mobile ad hoc network (MANET), providing insight for network defenders to understand how much information an adversary could discern about a target network. Using this topology inference strategy, a network eavesdropper collects frame emission start- and end-times and uses these to detect the presence of link layer acknowledgements between devices and ultimately constructs a network topology. We show how the performance of this simple strategy varies as a function of the amount of data collected by the eavesdropper over time, the size of the target network, the speed of the nodes, and the nodes' data generation rate. We derive analytical results that allow for the rapid computation of expected true positive rate and false positive rate for topology inference in a MANET; these are compared against simulation results. The analytical results are used to derive a sensible window of observation over which to perform inference, with guidance on when to discard stale data. The results are also used to recommend strategies for network defenders to frustrate the performance of an adversary's network inference.

Keywords: Network topology inference · Mobile ad hoc networks (MANET) · Traffic analysis · Cyber analytics

1 Introduction

The reconnaissance phase, in which an adversary gathers information about its intended target, is typically one of the first steps in any cyber-attack. In performing reconnaissance against a wireless ad hoc network, an important consideration is deriving the logical network topology: identifying which nodes have direct communication links with one another. A logical topology is either assumed or acquired as a pre-condition in many cyber and electronic warfare (EW) attacks. For instance, a simple man-in-the-middle attack [1] requires knowledge of where in the network to insert or capture packets; targeted jamming attacks such as in [2] require a topology map to achieve the greatest effect. Even for attacks in which knowing the target topology is not a pre-condition, having the topology map can lead to a greater impact; for example, the well-known wormhole [3] or black hole attacks [4] are most effective when they target known weak points in the network. Whether assumed or acquired, a known logical topology is the basis for an effective attack.

L. Foschini and M. El Kamili (Eds.): ADHOCNETS 2020, LNICST 345, pp. 70–86, 2021.
https://doi.org/10.1007/978-3-030-67369-7_6

In this paper, we present a general approach to topology inference in a mobile ad hoc network (MANET) and we analyze its performance. We derive analytical estimates for link detection accuracy and error rate, and we validate these through simulation. We also explore the relationship between node velocity and node traffic generation rate, leading to recommendations for defensive techniques to frustrate an adversary's topology inference.

Many works examining topology discovery assume that it is possible to probe the network in some fashion. For instance, [5] and [6] present techniques to infer a network topology using ICMP probes, where certain nodes in the network discard or ignore ICMP messages and others respond unreliably. While [5] and [6] are intended for mapping enterprise networks, these methods could also be effective in a wireless MANET in the case where an adversary had a packet injection capability but was not able to see the entire network from a single location. In [7], the difficulty of relying on ICMP is acknowledged, and another technique is introduced wherein non-ICMP probe messages are sent and the timing of the probes is measured to infer a topology; once again, however, the method assumes that it is possible to inject probes to which the network will respond. The M-iTop algorithm [8] adopts a different strategy that infers a network topology by vastly overestimating the number of components in the network and then paring them down through careful merging by relying on observed messages from known nodes within the network.

In unencrypted wireless networks using known communication protocols, it is generally understood that a global eavesdropper could determine the nearest-neighbour network topology without any probes simply by observing the link layer (or MAC) source and destination addresses for all packets in the network, inferring that source and destination nodes share a common link. Previous work in [9] identifies traffic flows in a MANET where the network is encrypted above the link layer and the MAC is obfuscated according to an anonymizing routing protocol such as in [10]. In this case, two-way communication between nodes over a short time period can still be used to perform one-hop topology inference based on short-term obfuscated identifiers (not fixed MAC addresses per se). End-to-end flow inference in a MANET is also explored in [11], which examines the correlation of packet timing and size across all nodes in the network to infer the presence of flows between certain parties; although not specifically focused on topology inference, it is a reasonable step to infer the network topology as well using this technique. While the technique in [11] shows promise, its performance is quite sensitive to tuning parameters and does not appear to be robust in the face of anonymizing routing protocols. In the case where encryption is extended to the link layer such that the MAC itself is encrypted, simple source-destination matching analyses are no longer possible.

This paper expands on a previous study [12], which assumes that an adversary detects direct links (i.e., direct point-to-point connections) between nodes in the network by observing the timing of link-layer messaging between the nodes. Specifically, an adversary observes acknowledgements (ACKs) sent in response to regular packet data and infers the existence of links between senders and receivers. By design, a link-layer ACK is typically an immediate short response following a link-layer data message between neighbouring nodes. In principal, these ACKs could be detected through packet timing analysis alone, even when encrypted, and without requiring access to headers or

payloads (indeed, the authors of [13] state without proof that a static network topology can be inferred in this manner). However, as we observed in [12], even for static networks using a simple ACK-based topology inference method, it is not evident *a priori* how much data must be observed to ensure the inferred network topology is accurate; nor is it evident what impact network size, network node range, and traffic flow frequency can have on that accuracy.

Our focus in this paper moves beyond static inference and examines a MANET with a continually changing topology. Our aim is to characterize the limitations and accuracy of such an ACK-based inference scheme, in which observed links become stale due to mobility-induced topology changes. These topology changes present a challenge to the accuracy of the inference scheme when one attempts to determine which links are still "fresh".

In this paper we make the following contributions:

1) We derive an analytical estimate of the accuracy of ACK-based topology inference in a MANET;
2) We provide supporting simulations that demonstrate the accuracy of the analytical estimate and give an intuitive sense of the roles of node velocity, node range, traffic flow frequency, and network size in performing topology inference;
3) We propose a method to select a sensible observation window size over which to perform inference (i.e., to decide when ACK messages are no longer "fresh" enough to be used to infer links); and
4) We offer guidance on how nodes in a network can frustrate the accuracy of the ACK-based traffic analysis by shaping the timing of their emissions as a function of the nodes' velocities.

The rest of the paper is organized as follows. Section 2 introduces the network and mobility model, along with basic assumptions and notation. In Sect. 3, we derive formulas that describe the correct detection rate and false detection rate for an ACK-based topology inference scheme. Simulations in Sect. 4 validate the derived formulas and show how performance varies with network size, node range, traffic load, and node velocity; this section also includes methods to select a window size for inference, and suggestions for network defenders. Section 5 summarizes our contributions with a brief conclusion.

2 Model, Assumptions and Notation

Following the notation in [14] (pp. 811–812), we represent a MANET of n nodes as a graph $G = (V, E)$, which describes a set of $V = \{1, 2, \ldots, n\}$ vertices representing the mobile nodes connected by edges $E \in V \times V$. We adopt the well-known random geometric graph (RGG) model (documented extensively by Penrose [15]) where we assume that nodes are positioned uniformly at random in a square area of dimensions X by X meters; any two nodes i and j are considered "neighbours" with a link between them (and hence an edge in the graph G) if the distance between the nodes, d_{ij}, is less than the network transmission range in meters, R. A route exists between two nodes if there exists a set of edges in E that connect the nodes through a set of vertices in V.

Nodes in the network generate traffic randomly according to a Poisson process at an average rate of λ packets per second; for each packet generated, a node selects a (routable) destination at random and transmits the packet over a potentially multi-hop route, where the average number of hops in the network is denoted by h.

Nodes move about from their initial positions following a random direction mobility model ([14], Chapter 7), where for mathematical tractability in Sect. 3 we assume all nodes travel at the same speed, v, and bounce off the boundary of the square area without pausing. While the set of nodes V never changes, the edges (i.e., the links, E) in G change over time as nodes move around in the square. Adopting a discrete-time model, we use the notation G_k to refer to the graph of the nodes in the network at time step k. Furthermore, the $n \times n$ adjacency matrix A_k is derived from G_k and describes the presence or absence of links between nodes; A_k consists of entries $a_{ij}[k] \in \{0,1\}$ for each pair of vertices (i, j) such that an edge exists between i and j if $a_{ij}[k] = 1$ and does not exist otherwise. We assume that links are symmetric such that $a_{ij}[k] = a_{ji}[k]$ for all (i, j).

In performing topology inference, we are interested in finding estimates $\hat{a}_{ij}[k]$ for each value of $a_{ij}[k]$ in A_k. If $a_{ij}[k] = \hat{a}_{ij}[k]$, then our inference about an edge between i and j is correct; otherwise it is incorrect. Using standard confusion matrix formulas [16], we can write the true positive rate (TPR) for the network-wide topology inference at time instant k as

$$TPR_k = \frac{1}{L_k} \sum\nolimits_{i \in V} \sum\nolimits_{j \in V | j<i} a_{ij}[k] \cdot \hat{a}_{ij}[k], \tag{1}$$

where L_k is the total number of edges in G_k; the TPR captures the percentage of links that have been correctly identified by a set of topology inference estimates $\hat{a}_{ij}[k]$ over all i and j. Likewise, the false positive rate (FPR) describes the percentage of non-links that have been falsely is identified as links and is computed using

$$FPR_k = \frac{1}{L_{max} - L_k} \sum\nolimits_{i \in V} \sum\nolimits_{j \in V | j<i} \left(1 - a_{ij}[k]\right) \cdot \hat{a}_{ij}[k], \tag{2}$$

where $L_{max} = n \cdot (n - 1)/2$ is the maximum possible number of links in a network of n nodes, and $L_{max} - L_k$ is the number of non-links in the network at instant k. Note that throughout the paper, we often drop the subscript k denoting the time step on TPR and FPR, though context should make it clear.

In modeling ACK-based topology inference, we assume a global eavesdropper that can observe the timing of all emissions from all nodes in the network and is able to distinguish these emissions from one another. We assume each unicast communication between any two adjacent nodes produces an identifiable ACK response, allowing the eavesdropper to infer the presence of the link. Note that the intent of our analysis here is not to determine how well we can identify ACKs, but to characterize the performance and limitations of a link-based inference scheme in a mobile network with a changing topology. A key realization from this last point is that while our work in this paper focuses on an ACK-based topology inference scheme, the results would apply equally well to any link-based topology inference scheme in which the transmission of a unicast packet allowed the eavesdropper to reliably infer the presence of a link between the source and destination of the packet.

3 Analytical Characterization of Topology Inference

Consider a MANET with n nodes represented by a random geometric graph with n vertices, where each node moves in an independently selected random direction with velocity v and generates traffic according to a Poisson process. In this section, we derive analytical expressions for the expected true positive and false positive rate of a link-based topology inference scheme operating in such a network. We introduce the notation T_k and F_k to represent the expected number of links correctly identified and falsely identified, respectively, at time step k. Thus, the expected true positive and false positive rates are expressed as $\text{TPR}_k = T_k/L$ and $\text{FPR}_k = F_k/(L_{max} - L)$, where L is the average number of links in the network and $L_{max} = n(n-1)/2$ and corresponds to the maximum possible number of links in the network.

Without loss of generality, consider the case where the duration of the time step from time k to $k+1$ is such that on average a single node in the network generates exactly one packet; this amounts to effectively choosing a time step size of $\Delta k = 1/\lambda n$. On average, a single unicast packet traverses h hops from source to destination; successfully observing ACK messages for each hop produces evidence for h links in the network arising from this unicast message. From time step k to $k+1$, we expect the average number of correctly identified links, T_k, to increase by some amount in response to this new evidence. In the absence of mobility, at time step k we would expect that some fraction of the h newly observed links are already known; in fact, the average number of observed links that are already known is simply h scaled by the expected true positive rate in the previous instant, $\text{TPR}_k = T_k/L$. Thus, for a static network we can write

$$T_{k+1} = T_k + h\left(1 - \frac{T_k}{L}\right). \tag{3}$$

Now, in a network with mobility, from time instant k to $k+1$, the nodes will have moved some distance, resulting in some number of link breakages (denoted here by L_{break}) and some number of new links (denoted here by L_{new}). Similar to the discussion above, we would expect that some fraction of the link breakages would have been previously (correctly) identified as true links, and this fraction would be given by the true positive rate in the previous time step, TPR_k; the fact that they are no longer links thus reduces the number of correctly observed links, T_{k+1}, by $\text{E}\{L_{break}\} \cdot (T_k/L)$, where $\text{E}\{L_{break}\}$ is the average number of link breakages expected over one time step.

Furthermore, some of the new links generated through mobility would have been previously (incorrectly) identified as links due to false detection, with this fraction of links given by the false-positive rate $F_k/(L_{max} - L)$. The result is that these newly formed links will now count among correct links, increasing T_{k+1} by $\text{E}\{L_{new}\} \cdot F_k/(L_{max} - L)$. Finally, as noted in [17], in a spatially constrained MANET with our constant-velocity mobility model we expect link generation and link breakage rates to be equal (on average), leading to $L_\Delta = \text{E}\{L_{break}\} = \text{E}\{L_{new}\}$, where L_Δ represents the average number of links broken (and created) in time step k to $k+1$. Incorporating the effects of mobility into Eq. (3) for T_{k+1} yields

$$T_{k+1} = T_k + h\left(1 - \frac{T_k}{L}\right) + L_\Delta\left(\frac{F_k}{L_{max} - L} - \frac{T_k}{L}\right). \tag{4}$$

Developing an expression for F_{k+1} follows in a similar fashion. Assuming links are detected accurately through ACK observations, the source of errors in our topology inference will arise primarily from out-of-date (stale) links that have broken due to node mobility. During time step k to $k + 1$ there will be $\mathrm{E}\{L_{break}\}$ links broken and $\mathrm{E}\{L_{new}\}$ links generated on average (where $\mathrm{E}\{L_{break}\} = \mathrm{E}\{L_{new}\} = L_{\Delta}$, as before). Some fraction of the links that break will have been previously correctly inferred, with this fraction given by TPR_k; these will now be in error, thus increasing the number of false positives, F_{k+1}, by $\mathrm{E}\{L_{break}\}\cdot(T_k/L)$. Likewise, some fraction of the new links would have been previously incorrectly inferred, as given by FPR_k; these will now be correct, thus decreasing F_{k+1} by $\mathrm{E}\{L_{new}\}\cdot F_k/(L_{max} - L)$. This results in an expression for F_{k+1} given by

$$F_{k+1} = F_k - L_{\Delta}\left(\frac{F_k}{L_{max} - L} - \frac{T_k}{L}\right). \tag{5}$$

An expression for L_{Δ} can be found by adapting the work in [18], in which it is shown that the link break and generation rates in a MANET with constant-velocity nodes are given by $(8/\pi)\rho R v$, where ρ is node density. This expression provides the average link break and generation rate per unit time as seen from the point of view of single node. To be consistent with our time step size and notation, we scale this expression by a factor of $(1/n\lambda)$ seconds/message. Additionally, we are interested in the link break/generation rate of the network, not just that seen by a single node—this requires that we scale the expression by a factor of $(n/2)$, where the factor of 2 arises since for all nodes in the network any link break/generation is viewed by the two nodes forming the link and we only want to count these once. With these modifications, we obtain the expression

$$L_{\Delta} = \frac{4}{\pi^2}\cdot\frac{N_{av}}{R}\cdot\frac{v}{\lambda}, \tag{6}$$

where we have also made the substitution $\rho = N_{av}/(\pi R^2)$, with N_{av} denoting the average number of neighbours seen by nodes in the network.

The expressions in (4) and (5) for T_{k+1} and F_{k+1} comprise a system of coupled first-order difference equations. Armed with Eqs. (4), (5) and (6), we now have analytical expressions to generate true positive and false positive rates for ACK-based topology inference in a MANET, given its size (n), node velocity (v), range (R), and traffic generation rate (λ). These expressions can be evaluated rapidly up to any desired time step, k, using simple programming, with initial conditions of $T_0 = F_0 = 0$. In evaluating (4), (5) and (6), values of N_{av} can be computed using the expression in [19], and the average number of links, L, is found using $L = \lambda N_{av}/2$. Values for the average number of hops, h, can be computed from simple network simulations, or using estimates such as those in [20].

Note that Eqs. (4), (5) and (6) are not specifically reliant on a particular choice of MANET routing protocol such as AODV (ad hoc distance vector) routing or OLSR (optimum link state routing). However, the average number of hops in the network can be influenced by the choice of such protocols; for instance if all routes were (unwisely) chosen to maximize the number of hops per route, this would clearly yield a different value for h than choosing to minimize the number of hops. In this paper, as discussed in

more detail in Sect. 4, our simulations use shortest-path routes; thus in evaluating TPR and FPR, above, we use an estimate for h based on shortest-path routing [20]. Of interest is that in static networks, we observed in [12] that a subset of links naturally form a part of multiple routes, making them more likely to be re-used as time goes on by a factor measured at roughly $(1 + (h - 1)T_k/L)^2$. Were a different routing protocol selected that did not explicitly minimize hops, a different estimate for h would be required, but this is beyond the scope of this paper.

We show in Sect. 4 that our expressions for expected TPR and FPR generate results that align very closely with network simulation.

4 Simulations and Analytical Results

We completed a series of simulations in MATLAB, where we considered the effect of varying network size, node range, node speed and traffic generation rate on the performance of an ACK-based topology inference scheme. In selecting parameters for our simulations, we considered small networks representative of military tactical network deployments. It is of interest for users of such networks to understand how much information an adversary might infer about their network topology. Simulation parameters are summarized in Table 1 and discussed below.

Table 1. Simulation parameters for MANET scenarios.

Simulation parameter	Values
Area of operation	Square area, 3000 m × 3000 m
Number of nodes (n)	8, 16, 24
Radio transmission range (R)	750 m, 1125 m, 1500 m
Node speed (v)	1 m/s, 3 m/s, 10 m/s
Average transmission rate per node (λ)	0.5 packets/s, 1 packet/s, 2 packets/s

A typical tactical network can be structured around the basic sub-unit of an infantry Section, which consists of approximately 8-10 nodes. Sections can operate independently or in groups of up to three, where three Sections are organized as a Platoon (approximately 24 nodes). Our simulations consider networks of sizes 8, 16, and 24 nodes, representing a single Section, two Sections operating together, and a Platoon.

Following guidance in [21], we consider a scenario where nodes travel in a square area of operation of size 3000 m × 3000 m. We assume that nodes have transmission ranges between 750 m to 1500 m, in line with performance expectations for existing high-bandwidth multi-hop radios. For simplicity of the model, all nodes move at the same speed, travelling at speeds of either 1 m/s (walking speed), 3 m/s (jogging speed), or 10 m/s (moderate vehicle speed). Each node randomly generates packets at a rate determined by a Poisson process with a mean, λ, of 0.5, 1, or 2 packets/s.

In conducting a single simulation run, we select values for each of n, R, v, and λ from Table 1. Nodes are placed uniformly at random in the area of operations and move at

speed v, changing directions when they encounter the edges of the area. Nodes generate traffic, which is transmitted via a shortest-path multi-hop route to intended (randomly selected) destinations.

During the course of a single simulation run, an ideal eavesdropper collects the timing of packets for all network transmissions and estimates the existence of ACKs by comparing the timestamp of each newly observed frame to the timestamp of the previously observed frame. If the difference between the two frames is less than a given threshold, the newly observed frame is deemed to be an ACK to the previous frame, indicating a link between the sources of the ACK and the previous frame. ACKs are assumed to be generated sufficiently rapidly (and are sufficiently short) that the time difference between a frame and its corresponding ACK is small enough to ensure correct detection of all ACKs.

Inferred links in a single run are compared against the known instantaneous ground truth of the network topology at every time step k, allowing for the computation of single-run TPR and FPR as functions of k. Average TPR and FPR are computed by repeating each single-run simulation 100 times (with new randomized node placement each time) and averaging all the single-run TPR and FPR values at each time step.

In Sect. 4.1, our plots examine the expected TPR and FPR as a function of k, without an attempt to drop old (stale) information; this is done to get a clear understanding of how correct and incorrect inference evolves over time. In Sect. 4.2 we explore how to select an appropriate time window to focus only on recently observed ACKs to limit the proliferation of inference errors due to stale information.

4.1 Effect of Range, Network Size, Node Speed, and Traffic Rate on Inference

To explore the effect of range, size, speed, and traffic generation rate on topology inference, we vary each of R, n, v and λ individually and plot the resulting average TPR and FPR against time step k (where to compare against estimates in (4) and (5), k is the time for the network to generate one new message, on average).

Figure 1 shows the effect of different transmission ranges, R, under mobility. We see that for lower values of R, the TPR increases more rapidly, while the FPR increases less rapidly than for higher R values. These results are intuitively satisfying, since a network with smaller R will have more hops on average between any source-destination pair; thus, a single source transmission is likely to hop more times through the network, revealing more links, leading to a higher TPR. Another factor that leads a smaller R to yield a higher TPR is that a smaller R results in a network with fewer links to discover, meaning that each transmitted message in the network provides proportionally more information. Likewise, with respect to false positives, a smaller R will result in a larger number of "non-links" in the network[1]; thus a single erroneously detected link will constitute a smaller fraction of the total non-links in the network, leading to a smaller FPR.

Figure 1 also shows agreement between the simulation results—shown as solid and dashed curves—and our analytical predictions derived from Eqs. (4) and (5), shown as dotted curves.

[1] A non-link in this context is a pair of nodes for which there is no direct 1-hop connection. There are, on average, $L_{max} - L$ non-links in the network.

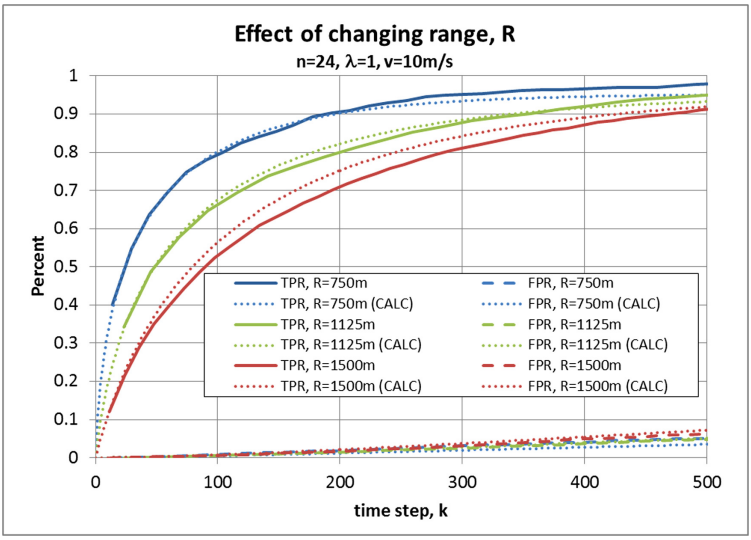

Fig. 1. Effect of transmission range, R, on the average true positive rate (TPR) and false positive rate (FPR) for MANET with $n = 24$, $v = 10$ m/s, $\lambda = 1$ packet/s.

The effect of network size on topology inference is shown in Fig. 2, which plots TPR and FPR for values of $n = 8$, 16, and 24. We observe that in a network with fewer nodes, there is a more rapid increase in both TPR and FPR compared to a larger network. The intuition behind TPR increasing faster for a smaller network is similar to what was described above for a network with lower range: the smaller network has comparatively fewer links to discover, meaning that the percentage of correctly inferred links is larger for each observed ACK. For FPR, however, the case is different here since a smaller network also has fewer potential non-links—this means that every error in inference (though infrequent) has a greater impact on the FPR compared to a larger network. Once again, we see that our simulations agree quite well with the analytical estimates in predicting the effect of network size on inference.

Figures 3 and 4 examine the effects of node velocity and data transmission rate on topology inference. A higher node velocity results in a more rapid increase in FPR; this is expected since faster nodes result in a more rapidly changing topology, meaning that links are broken more frequently—this in turn means that previously observed links become errors after a shorter period of time. An opposite phenomenon is observed for data transmission rate: a faster transmission rate means that for the same number of observed transmissions, k (with on average one new message occurring per time step k) we observe relatively fewer errors. This occurs since more transmissions occur in a shorter period of time, meaning there is less time for them to become stale. Note that this effect on FPR would be less apparent were we to scale the x-axis as a function of time as opposed to the discrete time step k (where we multiply k by $1/n\lambda$ to obtain time in seconds, as discussed in Sect. 3); instead we would observe a separation in TPR curves such that TPR rises more rapidly for larger values of λ.

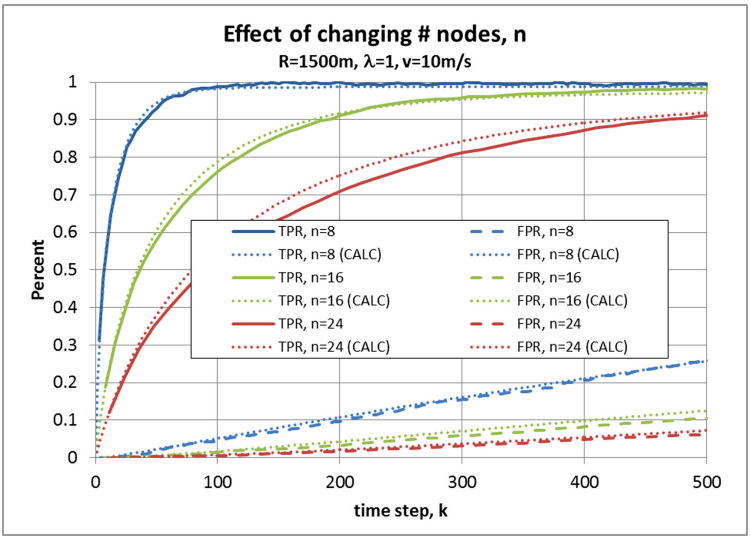

Fig. 2. Effect of network size, n, on the average true positive rate (TPR) and false positive rate (FPR) for MANET with $R = 1500$ m, $v = 10$ m/s, $\lambda = 1$ packet/s.

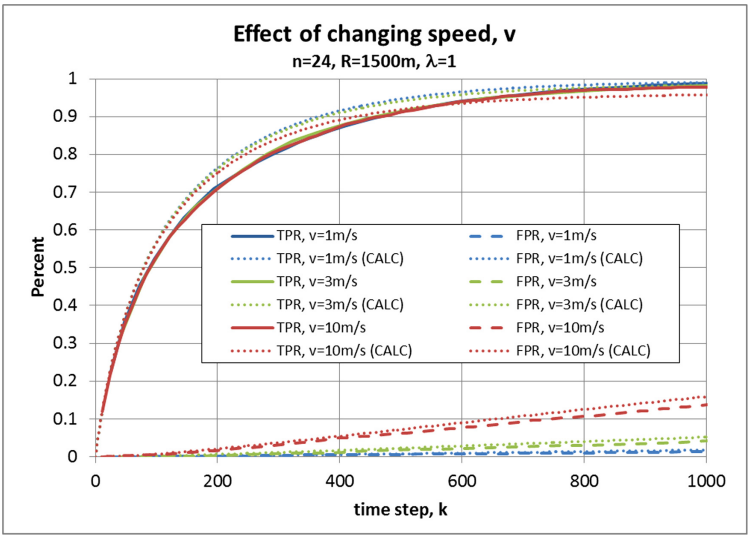

Fig. 3. Effect of node speed, v, on the average true positive rate (TPR) and false positive rate (FPR) for MANET with $n = 24$, $R = 1500$ m, $\lambda = 1$ packet/s.

Figures 3 and 4 suggest that node velocity and data transmission rate are important and related network metrics when it comes to characterizing the performance of topology inference. In fact, from Eq. (6), we observe that the rate of link breakage (and creation) in the network is a function of the ratio v/λ. We define a new term, $\gamma = v/\lambda$ with units

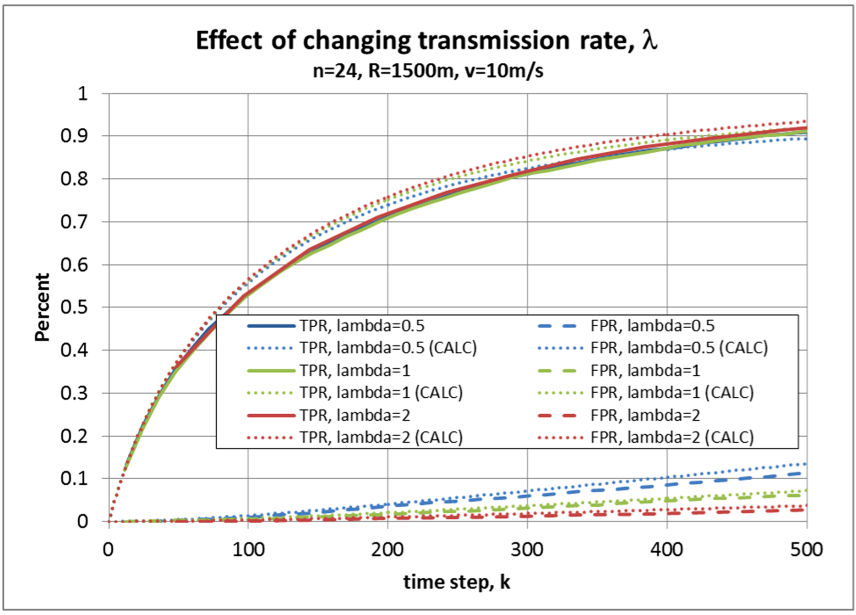

Fig. 4. Effect of packet transmission rate, λ, on the average true positive rate (TPR) and false positive rate (FPR) for MANET with $n = 24$, $R = 1500$ m, $v = 10$ m/s.

of meters/packet—ultimately this refers to the expected distance in meters that a node would travel before generating one packet. Based on (4), (5) and (6), we would expect that TPR and FPR (as functions of k) should depend only on the ratio γ, and not on the individual values of v and λ. Figure 5 plots simulation results for TPR and FPR, showing that as long as v and λ change in sync and maintain a constant ratio, γ, our performance is unchanged.

4.2 Selecting Window of Observation for Topology Inference

In Sect. 4.1, the expected true positive rate and false positive rate were plotted assuming all inference data is retained for the duration of the simulation. In practice, since the network is mobile we would expect that older data would become less relevant; indeed, in our model the primary source of false positives arises from links that were (correctly) inferred in the past and which no longer exist.

One potential means of reducing errors is to simply use a sliding window, such that topology inference is only performed using data that is newer than some threshold (i.e., fits within the time window)[2]. We would expect that by limiting our observations to a sliding window of m time steps the expected TPR and FPR would reach steady state

[2] In this work, we consider fixed time windows where all data in the window is treated equally until its age exceeds the window size. An alternative for future work would be to include a decay factor where data has less impact on an inference decision as it ages, as opposed to a strict step function like with our window.

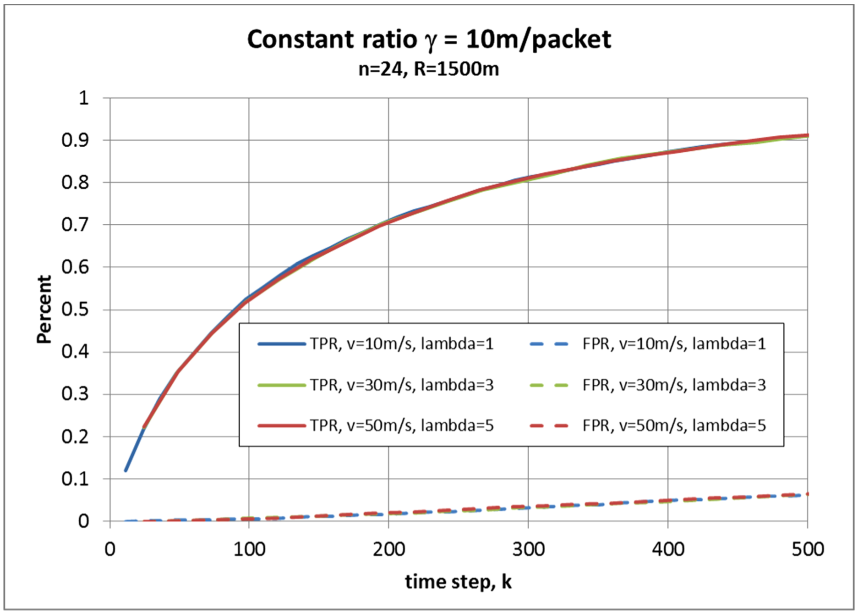

Fig. 5. TPR and FPR in a MANET with constant $\gamma = 10$ m/packet, with $n = 24$ and $R = 1500$ m.

given by the values of TPR[m] and FPR[m] from our plots in Sect. 4.1. In fact, this is precisely what we observe in Fig. 6, where we apply several window sizes to the data in our simulation.

A larger window size results in a higher expected TPR, but also results in a higher expected FPR. The question is how to select an appropriate window size when performing topology inference. Ultimately, the answer depends upon the particular goals of the adversary performing the inference and what is deemed an acceptable[3] trade-off between TPR and FPR. One well-known method to choose a balance between TPR and FPR in diagnostic or binary decision problems is to maximize Youden's J statistic (sometimes called the informedness)[4], where $J_k = \text{TPR}_k - \text{FPR}_k$ [22]. Figure 7 shows a plot of J, TPR and FPR for a network with $n = 24$, $R = 1500$ m, $v = 10$ m/s and $\lambda = 1$ packet/s. Finding the value at which J_k is maximized is a simple matter of finding the value for k at which $J_{k+1} - J_k = 0$; this is shown by the highlighted area in Fig. 7.

In Fig. 8 we plot the J statistic for a range of values of $\gamma = v/\lambda$. We use Eqs. (4) and (5) to compute the TPR and FPR curves that inform the J statistic in this case. The recommended window size as a function of γ is shown in Fig. 9. We observe that the

[3] We note that the relationship between TPR and FPR for binary classification problems often involves plotting TPR versus FPR as a receiver operating characteristic (ROC) curve; however, for our purposes it is more informative to observe these values plotted against a common axis.

[4] Youden's J statistic is typically expressed as $J =$ sensitivity + specificity – 1. The sensitivity of a measurement in statistics is equal to the TPR, and the specificity is (1-FPR). These substitutions lead to $J =$ TPR – FPR. The intuition behind the J statistic is that it is the point on the ROC curve furthest from chance.

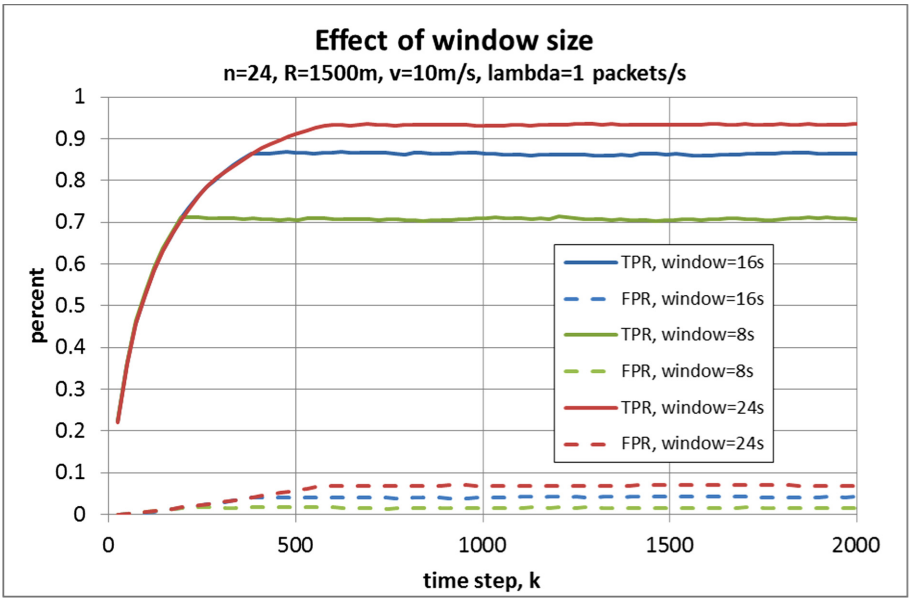

Fig. 6. True positive rate (TPR) and false positive rate (FPR) for topology inference based on data constrained to a sliding window.

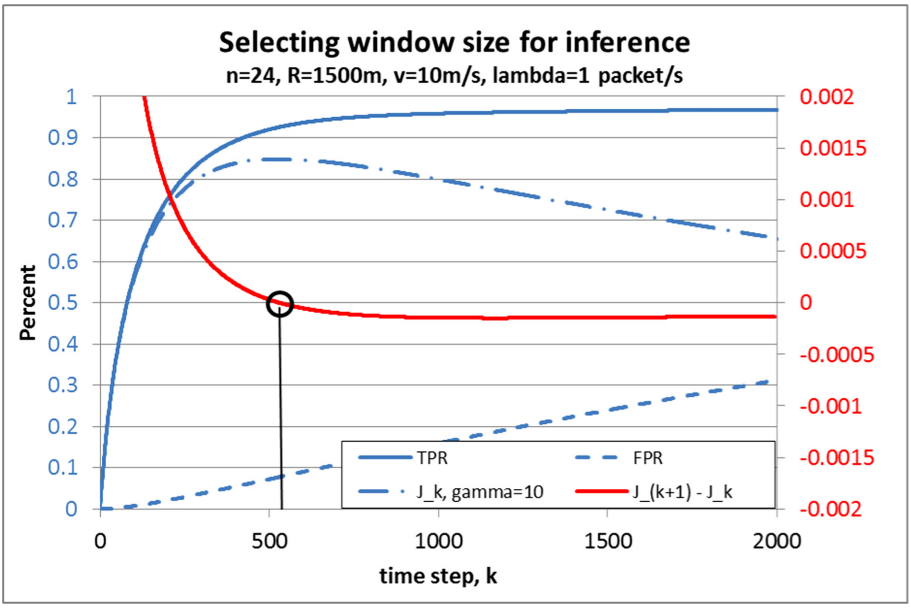

Fig. 7. Selecting a window size for topology inference by maximizing TPR - FPR.

window size decreases as γ increases. This result largely agrees with the intuition gained in Sect. 4.1. If the nodes are moving faster (larger v), then data will become stale more quickly and we should have a smaller window size k; similarly if the network produces data more slowly (smaller λ), we will accumulate less data before it gets stale and thus would have a smaller window size, k (where once again, k is the number of discrete time steps in the window and can be converted to time in seconds by scaling by $1/n\lambda$).

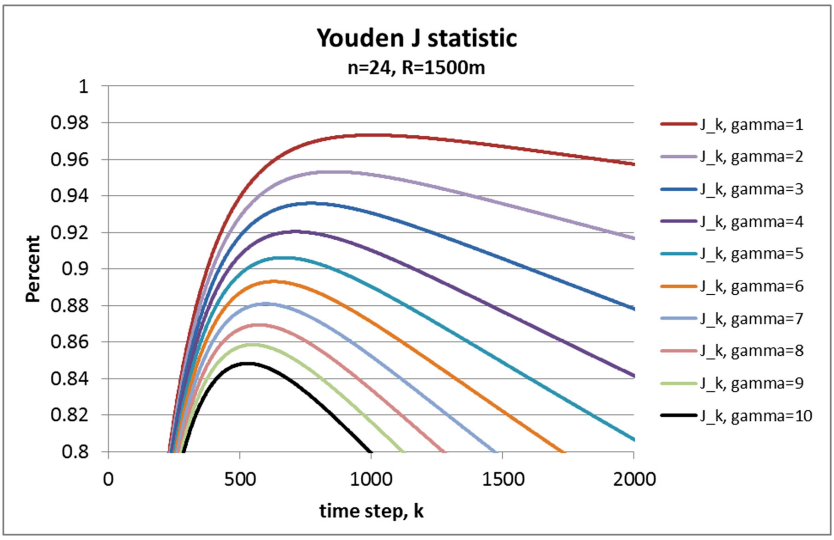

Fig. 8. The J statistic is shown for a range of values, $\gamma = v/\lambda$.

Figure 8 provides an interesting conclusion for users wishing to limit an adversary's success in performing network topology inference. We observe that the ultimate success rate of inference (as described by the J statistic) is lower for higher values of γ. Thus, where feasible, increasing node speed or decreasing data transmission rate (or both) will succeed in increasing the value γ, leading to poorer inference. While a network operator may have little control over node speeds, data rate is more fungible. A network operator with a desired set-point for γ could in theory adjust data rates, beacons, and flow control to keep the ratio γ within a desired range depending upon node speed and network dynamics.

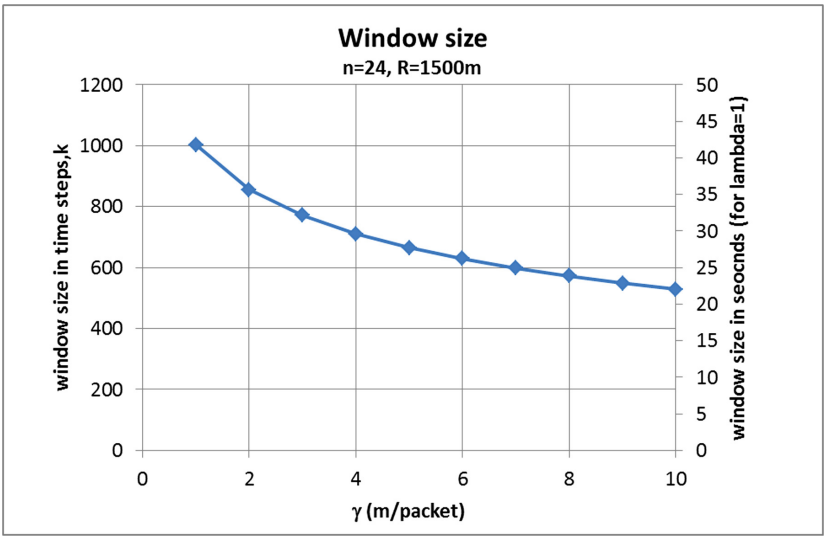

Fig. 9. Recommended window size for topology inference showing that window size decreases as a function of γ.

5 Conclusion

This paper characterized the performance of an ACK-based topology inference scheme for mobile ad hoc networks. We derived accurate analytical estimates that allow for rapid computation of the network-wide topology inference true positive rate and false positive rate, assuming constant velocity nodes in a MANET modeled as a random geometric graph. We validated our analytical estimates through extensive simulations; the simulations showed how topology inference is influenced by network size, node transmission range, node speed, and data transmission rate.

Based on our analysis, we identified valuable conclusions for network operators wishing to reduce an adversary's ability to perform topology inference: increasing the network size (i.e., the number of nodes in the network) where possible will require an adversary to collect more data over a longer period of time to identify all network links; increasing the ratio of node speed to data transmission rate will reduce the adversary's accuracy in topology inference—this can be accomplished by increasing node speed, decreasing transmission rate, or both.

Our analysis in this paper made a number of simplifying assumptions regarding the adversary's eavesdropping capability and the network dynamics. In future work, we plan to examine the effect of limited adversary range (as opposed to an ideal adversary that can listen to the entire network); heterogeneous network nodes that operate at different velocities and generate data at different rates; and how the timing data of higher-layer network protocol information (beyond link-layer ACKs) can be incorporated.

References

1. Vanhoef, M., Piessens, F.: Advanced WiFi attacks using commodity hardware. In: Proceedings of 2014 Annual Computer Security Applications Conference, New Orleans LA, USA (2014)
2. Eisen, J., Watson, S., Willink, T.: Location constrained jamming: surgical link removal using local graph partitioning. In: Proceedings of 2018 International Conference on Military Communications and Information Systems (ICMCIS), Warsaw, Poland (2018)
3. Hu, Y.C., Perrig, A., Johnson, D.B.: Packet leashes: a defense against wormhole attacks in wireless networks. In: Proceedings of 22nd Annual Joint Conference of the IEEE Computer and Communications Societies, San Francisco, USA (2003)
4. Puray, M., Palod, P.: Black-hole attack in MANET: a study. Int. J. Adv. Res. Comput. Eng. Technol. **5**(3) (2016)
5. Jin, X., Ken Yiu, W.P., Gary Chan, S.H., Wang, Y.: Network topology inference based on end-to-end measurements. IEEE J. Sel. Areas Commun. **24**(12), 2182–2195 (2006)
6. Gunes, M.H., Sarac, K.: Resolving anonymous routers in internet topology measurement studies. In: Proceedings of IEEE INFOCOM 2008, Phoenix AZ, USA (2008)
7. Malekzadeh, A., MacGregor, M.H.: Network topology inference from end-to-end unicast measurments. In: Proceedings of 27th International Conference on Advanced Information Networking and Applications Workshops, Barcelona, Spain (2013)
8. Silvestri, S., Holbert, B., Novotny, P., La Porta, T., Wolf, A., Swami, A.: Inferring network topologies in MANETs applied to service redeployment. In: Proceedings of 24th International Conference on Computer Communication and Networks (ICCCN), Las Vegas NV, USA (2015)
9. Liu, Y., Zhang, R., Shi, J., Zhang, Y.: Traffic inference in anonymous MANETs. In: Proceedings of 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Boston MA, USA (2010)
10. Zhang, Y., Liu, W., Lou, W.: Anonymous communications in mobile ad hoc networks. In: Proceedings of IEEE INFOCOM 2005, Miami FL, USA (2005)
11. Chang, H., Shan, H.: End-to-end flow inference of encrypted MANET. In: Proceedgins of IEEE 3rd International Conference on Information Science and Technology, Yangzhou, China (2013)
12. Brown, J.D., Salmanian, M., Willink, T.J.: Topology inference of multi-hop wireless networks. DRDC Scientific Report, DRDC-RDDC-2018-R300 (2019)
13. He, T., Wong, H.Y., Lee, K.: Traffic analysis in anonymous MANETs. In: Proceedings of IEEE Military Communications Conference (MILCOM), San Diego CA, USA (2008)
14. Roy, R.R.: Handbook of Mobile Ad Hoc Networks For Mobility Models. Springer, Boston (2011)
15. Penrose, M.D.: Random Geometric Graphs. Oxford University Press, Oxford (2003)
16. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)
17. Samar, P., Wicker, S.B.: On the behavior of communication links of a node in a multi-hop mobile environment. In: Proceedings of MobiHoc Conference, Tokyo, Japan (2004)
18. Cho, S., Hayes, J.P.: Impact of mobility on connection stability in ad hoc networks. In: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), New Orleans LA, USA (2005)
19. Bakhshi, B., Khorsandi, S.: Node connectivity analysis in multi-hop wireless networks. In: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), Sydney Australia (2010)
20. Younes, O., Thomas, N.: Analysis of the expected number of hops in mobile ad hoc networks with random waypoint mobility. Electron. Notes in Theoret. Comput. Sci. **275**, 143–158 (2011)

21. Chapman, B.J.: Bounding the force employment concept. Technical Memorandum, Defence R&D Canada Centre for Operation Research and Analysis (2009)
22. Powers, D.M.W.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J. Mach. Learn. Technol. **2**(1), 37–63 (2011)