

An Information Theoretic Approach for MANET Unlinkability Measure

Dijiang Huang and Yang Qin

Arizona State University, USA
{dijiang, yang.qin.1}@asu.edu

ABSTRACT

Measuring communication anonymity (e.g., unlinkability) of mobile wireless ad hoc networks is a critical but still unsolved problem. In this paper, we present a theoretic model for unlinkability measure in mobile ad hoc networks. In our approach, we consider evidence as measurements (e.g. packet count), which is used in this paper to build our unlinkability measuring models. Our approach is based on evidence theory, where the basic measuring component is “set”. We present theoretical models to evaluate end-to-end unlinkability measure, single path unlinkability measure, and system unlinkability measure. We present two techniques to show the quality of an unlinkability measure. Finally, we conduct simulation studies to illustrate the use of our unlinkability evaluation models.

1 Introduction

To measure anonymity performance of an anonymous communications, information theoretic models are prevalent among all existing unlinkability measuring models (such as [1]–[3]) due to its simplicity and broad applicability to many anonymous systems. Information theoretic models are built on probability theory and are suitable for statistical unlinkability analysis and evaluations [4]. However, in mobile wireless systems, probability theoretic models have some restrictions to measure anonymity performance based on fuzzy information. Therefore we propose a more general mathematical model for unlinkability evaluation based on evidence theory.

Probability theory is a branch of evidence theory, while evidence theory is, in turn, a branch of fuzzy measure theory. In evidence theory, given a universal set \mathbf{X} , the belief measure $Bel : \mathcal{P}(\mathbf{X}) \rightarrow [0, 1]$ denotes the lower bound of an unlinkability measure [5]:

$$Bel(\mathbf{V}) = \sum_{\mathbf{U} \subseteq \mathbf{V}} p(\mathbf{U}) = 1 - \sum_{\mathbf{U} \not\subseteq \mathbf{V}} p(\mathbf{U}), \quad (1)$$

where $p(\mathbf{U})$ is the basic probability assignment for the evidence given in \mathbf{U} . In (1), only if the evidence \mathbf{U} is a subset of \mathbf{V} , will it be counted as the evidence supporting the claim \mathbf{V} . Note that if \mathbf{V} is singleton, we have $Bel(\mathbf{V}) = p(\mathbf{V})$ where $p(\mathbf{V})$ is the probability assignment of the singleton \mathbf{V} and a belief measure is reduced to a probability measure. Compared to the probability theory, the basic element in evidence theory is a *set*. Every set is crisp and the issue is the likelihood of membership in each of these sets of an object whose characterization is imprecise and, possibly, fuzzy. This property perfectly fits into the wireless communication system.

In this paper, we firstly introduce evidence theory for unlinkability measure in mobile ad hoc networks. We present a systematic approach to demonstrate how evidence theory can be used for unlinkability measure. The proposed evidence-based solutions deal with set involvement inferences by using various evidence-based measurements, such as belief measure, plausibility measure, etc [6]. In particular, we propose a new concept “BP” unlinkability measure which is a compromised evaluation of two extreme evaluations, i.e., belief measure (lower bound) and plausibility measure (upper bound). In addition, we present theoretical models to evaluate end-to-end unlinkability measure, single path unlinkability measure, system unlinkability measure, amplification measure, and KL measure [7]. Finally, we present a comparative simulation studies to illustrate the use of the proposed models and conclude unlinkability findings. Our research aims to provide a more general methodology for unlinkability measure. As we described previously, evidence theory is the generalized version of probability theory. It can be reduced to existing probability theory-based measure. We hope our proposed research will build a general traffic analytical model to evaluate unlinkability in communication systems.

Our paper is arranged as follows: in section 2, we present the assumptions and definitions ; in section 3, we propose the evidence theory based unlinkability measure for mobile ad hoc networks; in section 4, we describe how to evaluate an unlinkability measure; in section 5, we perform a simulation study to estimate the unlinkability of an MANET using the proposed unlinkability measure; finally, we conclude our work in section 6.

2 Assumptions and Definitions

In this section, we present the assumptions and definitions to be used in the proposed unlinkability measure.

2.1 Assumptions

The following capabilities of wireless signal detectors are assumed: (a) they can detect, capture, and monitor the traffic transmitted within a wireless communication system, however they cannot decrypt the content of captured frames; (b) they are silent (passive) observers without injecting frames or interrupting frame transmissions; (c) they can locate wireless stations and trace their movements; (d) they can detect the wireless signal transmission power and transmission directions at any given location, i.e., they can locate the signal source; (e) the mobile devices’ hardware properties are known, thus the transmission range and physical carrier sensing range of a mobile device are known; (f) due to the resource

restriction in MANET, packet padding and delay are not used. Based on the above described abilities, the detectors can build the one-hop traffic matrices.

2.2 Definitions

We present a formal definition of anonymity based on the description in [8].

Definition 1 (Anonymity) *Anonymity is in the state of being not identifiable within an anonymity set \mathbf{V} ; the anonymity set is the set of all possible entities $v_k \in \mathbf{V}$ ($k = 1, \dots, N$).* ■

If we consider sending and receiving of messages as the items of interest (IOIs), anonymity may be defined as unlinkability of an IOI and any subject. More specifically, we can describe the anonymity of an IOI such that it is not linkable to any subject, and the anonymity of a subject as not being linkable to any IOI [8]. In a communication system, we evaluate the confidence (the level of assurance) to determine who communicates with whom (i.e. the source-destination relation of a transmission) as the unlinkability measure [4].

A set of entities is represented as \mathbf{V} . A communication relation $\mathcal{X}_{o \rightarrow d}$ is represented as a directional mapping from o to d , where $\mathcal{X}_{o \rightarrow d} \in \mathbf{V} \times \mathbf{V}$, $o, d \in \mathbf{V}$, o is the message source and d is the corresponding destination.

Definition 2 (Unlinkability) 1) *The Anonymity set is a set $\mathbf{V} \subseteq \mathbf{X}$ containing all potential acting entities that is bounded by a time interval Δt . A communication relation \mathcal{X} within Δt can be defined as $\mathcal{X}_{\Delta t, o \rightarrow d}$ where $o, d \in \mathbf{V}$. (In many places, we omit subscript Δt to simplify the notation).*

2) *The sender/receiver unlinkability is measured by the probability $p : o \rightarrow d$ of successfully identifying a transmission relation $\mathcal{X}_{o \rightarrow d}$. The probability p also represents the confidence on believing the relation $\mathcal{X}_{o \rightarrow d}$.*

3) *Perfect unlinkability: For a given pair of entities $(o, d) \in \mathbf{V}$, we can derive the perfect unlinkability measure for each communication relation \mathcal{X} as:*

$$H(\mathcal{X}) = -\log_2 p(\mathcal{X}) = \log_2 |\mathcal{X}|, \quad (2)$$

where $H(\mathcal{X})$ is the Hartley uncertainty measure [9], and $p(\mathcal{X}) = 1/|\mathcal{X}|$ is the probability to identify the communication relation for each communication pair, such as the sender o or the receiver d defined by $\mathcal{X}_{o \rightarrow d}$, from the anonymity set \mathbf{V} . If we only refer to sender unlinkability or receiver unlinkability, we can rewrite (2) as the follows:

$$H(\mathbf{V}) = -\log_2 p(\mathbf{V}) = \log_2 |\mathbf{V}|. \quad (3)$$

The equation (2) implies that the maximum unlinkability can be achieved when the communication system is in an indistinguishable state. In other words, we cannot derive the preference of one event over another. In (3), identifying a sender or a receiver is equivalent to identifying an entity from the anonymity set \mathbf{V} . Thus, achieving perfect unlinkability of a communication system is equivalent to random guessing.

If the communication party o is known, determining another party is equivalent to identifying the entity within the unlinkability set $\mathbf{V} \setminus \{o\}$. In this case, the size of relation universe \mathbf{X} is reduced to the size of unlinkability set $\mathbf{V} \setminus \{o\}$.

Definition 3 (Inclusive Set) *Set $\mathbf{V}_{\langle o, d \rangle} = \langle o, \mathbf{\Lambda}, d \rangle$ represents a set of ordered sets starting by entity o and ending by entity d , where $\mathbf{\Lambda} = \{\forall \lambda | \lambda \subseteq \mathcal{P}(\mathbf{V} \setminus \{o, d\})\}$, $\mathcal{P}(\mathbf{V})$ is the power set of \mathbf{V} and $\mathcal{P}(\mathbf{V} \setminus \{o, d\})$ is the power set of $\mathbf{V} \setminus \{o, d\}$. Thus, $\mathbf{V}_{\langle o, d \rangle}$ is a group of ordered sets that start from o and end by d . We define inclusive subset relation $\mathbf{u} \subseteq \mathbf{V}_{\langle o, d \rangle}$, if $\mathbf{u} = \{o, \mu, d\}$ and $\mu \subseteq \mathbf{\Lambda}$.* ■

In rest of presentations, we use \mathbf{V} to represent a set, $\mathbf{V}_{\langle o, d \rangle}$ to represent an inclusive set, and $\mathbf{V}(s)$ to represent an ordered set where the sequence is specified in s .

Definition 4 (Evidence Matrices) *We define the evidence matrices for a wireless network with N nodes as follows:*

$$\mathbf{W}|_{1 \times K} = \{\mathbf{W}^{\Delta t_k} | k = 1, \dots, K\},$$

where $\mathbf{W}^{\Delta t_k} = (w_{(i-j)}^{\Delta t_k})_{N \times N}$ is a $N \times N$ square evidence matrix during time interval Δt_k ; $w_{(i-j)}^{\Delta t_k}$ is the directional point-to-point evidence from node i to node j ($i \neq j$ and $i, j \in \mathbf{V}$) during the time interval Δt_k . $w_{(i-i)}^{\Delta t_k}$ denotes the amount of evidence originated from the node i and we have the following traffic constraint:

$$\forall j, \quad w_{(i-j)}^{\Delta t_k} \leq w_{(i-i)}^{\Delta t_k}, \quad \text{where } i \neq j.$$

The evidence $w_{(i-j)}^{\Delta t_k}$ can be accumulative such as packet counts (or volume), delay, etc. ■

Definition 5 (Communication Relation Matrix) *We define the communication relation matrix as:*

$$\mathbf{R}^{t_K} |_{f(\mathbf{W}|_{1 \times K})} = (r_{(i,j)}^{t_K})_{N \times N}, \quad (4)$$

where $r_{(i,j)}^{t_K}$ denotes the end-to-end maximum amount of evidence from node i to node j via all possible paths within the time period confined by $t_K = \sum_{k=1}^K \Delta t_k$. For example, in terms of traffic volume, $r_{(i,j)}^{t_K}$ represents the maximum actual and deduced data traffic transmitted from node i to node j based on the captured evidence $\mathbf{W}|_{1 \times K}$. \mathbf{R}^{t_K} is derived from a sequence of traffic matrices $\mathbf{W}|_{1 \times K}$ by using a transformation function f . The transformation function f can be versatile and it depends on the implementation of the communication model. ■

In evidence theory, for the universal set \mathbf{X} , the belief measure $Bel : \mathcal{P}(\mathbf{X}) \rightarrow [0, 1]$ denotes the lower bound of an unlinkability measure [5]:

$$Bel(\mathbf{V}) = \sum_{\mathbf{U} \subseteq \mathbf{V}} p(\mathbf{U}) = 1 - \sum_{\mathbf{U} \not\subseteq \mathbf{V}} p(\mathbf{U}), \quad (5)$$

where $p(\mathbf{U})$ is the basic probability assignment for the evidence given in \mathbf{U} . The *belief measure* in (5) denotes the lower

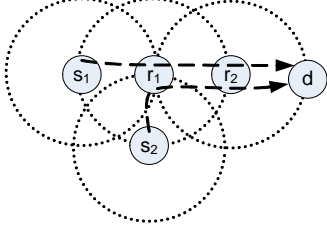


Figure 1: Plausibility Measure vs. Belief Measure.

bound of an unlinkability measure. In (5), only if evidence \mathbf{U} is a subset of \mathbf{V} , will it be counted as the evidence supporting the claim \mathbf{V} . This is intuitive since a packet delivering path is determined in an ordered set \mathbf{V} and it is supported by the evidence of its sub paths.

In evidence theory (a.k.a., *Dempster-Shafer theory* [5]), the dual function of belief measure function is called *plausibility measure* $Pl(\mathbf{V}) : \mathcal{P}(\mathbf{X}) \rightarrow [0, 1]$, which represents the upper bound of an unlinkability measure.

$$Pl(\mathbf{V}) = \sum_{\mathbf{U} \cap \mathbf{V} \neq \emptyset} p(\mathbf{U}) = 1 - \sum_{\mathbf{U} \cap \mathbf{V} = \emptyset} p(\mathbf{U}). \quad (6)$$

In (6), $Pl(\mathbf{V})$ represents the upper-bound of an unlinkability measure. As long as $\mathbf{U} \cap \mathbf{V} \neq \emptyset$, \mathbf{U} is considered as a supporting evidence of the claim \mathbf{V} . It is easy to derive the relation $Pl(\mathbf{V}) = 1 - Bel(\bar{\mathbf{V}})$ and $Pl(\mathbf{V}) \geq Bel(\mathbf{V}), \forall \mathbf{V} \in \mathcal{P}(\mathbf{X})$.

In general, the adversaries' evidence measuring abilities determine which measure is to be used. It also shows the confidence of adversaries on unlinkability measurements based on a given set of evidence. For example, if the adversary can determine the paths relation $\mathbf{U} \subseteq \mathbf{V}$, where $\forall \mathbf{U}, \mathbf{V} \in \mathcal{P}(\mathbf{X})$, the adversary has more confidence in using the belief measure. In Figure 1, the evidence supporting the transmission from s_2 to r_1 will also contribute to the evidence that supports the end-to-end communication from s_1 to d since $\langle s_2, r_1 \rangle \cap \langle s_1, r_1, r_2, d \rangle \neq \emptyset$. However, in belief measure this transmission evidence will not be counted since $\langle s_2, r_1 \rangle \not\subseteq \langle s_1, r_1, r_2, d \rangle$. In MANET, for example, if packet padding is not deployed, we can be more sure of multiple hop communication relations if several captured packets has the same size when table driven anonymous MANET routing schemes (such as [10]) are used.

In a MANET, it is difficult to determine that the unlinkability measure is *Bel* measure or is *Pl* measure. This is because evidence measure relies on the fuzziness of detected evidence. In other words, the traffic detection and evidence collection techniques determine whether the evaluation prefers either a *Bel* measure, a *Pl* measure, or something in between. Thus, we propose a general term *BP* measure that represents the unlinkability measure which is given as follows:

Definition 6 (BP Probability Assignments)

The evidence can be weighted or added using its accumulative form. Based on the construction in the Definition 5, we define the following notations:

$$r_{\langle o, * \rangle}^{t_K} = \sum_{\forall v \neq o} r_{\langle o, v \rangle}^{t_K}, \quad r_{\langle *, d \rangle}^{t_K} = \sum_{\forall u \neq d} r_{\langle u, d \rangle}^{t_K}, \quad r_{\langle *, * \rangle}^{t_K} = \sum_{u \neq v} r_{\langle u, v \rangle}^{t_K}.$$

We define the BP probability assignments for the originator and destination pair (o, d) during the time period t_K as follows:

$$\beta_r(\mathbf{V}_{\langle o, d \rangle}) = r_{\langle o, d \rangle} / r_{\langle o, * \rangle}, \quad \sum_{\forall d \neq o} \beta_r(\mathbf{V}_{\langle o, d \rangle}) = 1; \quad (7)$$

$$\beta_s(\mathbf{V}_{\langle o, d \rangle}) = r_{\langle o, d \rangle} / r_{\langle *, d \rangle}, \quad \sum_{\forall o \neq d} \beta_s(\mathbf{V}_{\langle o, d \rangle}) = 1; \quad (8)$$

$$\beta(\mathbf{V}_{\langle o, d \rangle}) = r_{\langle o, d \rangle} / r_{\langle *, * \rangle}, \quad \sum_{o \neq d} \beta(\mathbf{V}_{\langle o, d \rangle}) = 1. \quad (9)$$

■

$\mathbf{BP}_r = (\beta_r(\mathbf{V}_{\langle o, d \rangle}))_{N \times N}$, $\mathbf{BP}_s = (\beta_s(\mathbf{V}_{\langle o, d \rangle}))_{N \times N}$, and $\mathbf{BP} = (\beta(\mathbf{V}_{\langle o, d \rangle}))_{N \times N}$ are unlinkability probability assignment matrices of receiver, sender, and system, respectively. In the Definition 6, $\mathbf{V}_{\langle o, d \rangle}$ represents an inclusive set (defined in the Definition 3) which includes all possible subsets $\mathbf{u} \subseteq \mathbf{V}_{\langle o, d \rangle}$. This requirement is realized by the f function in Definition 5. In Section 3, we will present the construction of f function.

Definition 7 (Body of Evidence) For the BP probability assignment $\beta(\mathbf{V}_{\langle o, d \rangle}) \neq 0$, the set $\mathbf{V}_{\langle o, d \rangle} \subset \mathcal{P}(\mathbf{X})$ is called a focal element denoted by \mathcal{F} ; $\mathcal{B} = \{\mathcal{F}, \beta\}$ is called a body of evidence, which represents the set (i.e., the set includes all BP probability assignments) of all focal elements induced by β . ■

3 Evidence Theory Based Unlinkability Measure for Mobile Ad Hoc Networks

To hide communication relations, traffic shaping technologies have been deployed and they aim to the following goals: (a) evenly distributed traffic and constant rate management, (b) the end-to-end packet delivery timing information is not detectable or not distinguishable from multiple flows, and (c) both traffic volume and timing are non-correlatable. In Internet, the traffic padding and timing control are used to achieve (a) and (b), respectively. Due to the resource constraints, these techniques are not commonly adopted by anonymous MANET communications [4]. In order to illustrate our proposed unlinkability analytic model, we will present a packet counting-based (or traffic volume) technique to measure three typical unlinkability properties, i.e., the unlinkability for end-to-end communication relations, packet forwarding paths, and the whole MANET communication system.

3.1 End-to-end Unlinkability Measure

In [4], the author described the packet-counting traffic analysis attack. In such an attack, adversaries count the number of transmitted packets between each pair of mobile entities as the communication relation evidence. We present a generalized framework to analyze such an attack. Traffic volume (or packet count) anonymity metric is defined as follows: within a given time period Δt , the adversary captures the traffic volume matrix $\mathbf{W} = (w_{(i-j)})_{N \times N}$, where

N is the size of the network, $w_{(i-j)}$ is the point-to-point traffic captured from node i to node j . Based on the Definition 5, we can derive the communication relation matrix: $\mathbf{R}^{t\kappa}|_f(\mathbf{W}|_{1 \times K}) = (r_{(i,j)})_{N \times N}$. The transformation function f is given as follows:

$f(\mathbf{W} _{1 \times K})$ $\mathbf{R}^{t_1} = \mathbf{W}_1;$ for $i=1$ to $K-1$ $\mathbf{R}^{t_{i+1}} = \mathbf{R}^{t_i} + \mathbf{W}_{i+1} + g(\mathbf{W}_i \cdot \mathbf{W}_{i+1});$

Function g includes the following consecutive operations:

- (i) Compute the product of two matrices $\mathbf{W}'_i = \mathbf{W}_i \cdot \mathbf{W}_{i+1}$. The product of two one-hop traffic matrices deduce the two-hop traffic transformation matrix.
- (ii) Set the diagonal values of \mathbf{W}'_i to 0. The diagonal of a transformation matrix is the loop back traffic deduction (i.e., the same packet returns to the sender) and it should be removed.
- (iii) A timing threshold \mathcal{T} and a hop-count threshold \mathcal{H} are used to filter out packets that travel too long and too far in \mathbf{W}'_i and then return \mathbf{W}'_i .

After applying the transformation function f on the evidence matrices $\mathbf{W}_{1 \times K}$, we can derive the end-to-end traffic relation matrix $\mathbf{R}^{t\kappa}$. Using the *BP* probability assignments present in Definition 6, we can derive the communication relation probability for any end-to-end communication relation $r_{(i,j)}$.

3.2 Unlinkability Measure for A Path

Hartley addressed [9] that the amount of uncertainty associated with a set of alternatives can be measured by the amount of information needed to remove the uncertainty. Under set constructions, we can rewrite (3) for an ordered set $\mathbf{V}(s)$ as follows:

$$H(\mathbf{V}(s)) = \log_2 |\mathbf{V}(s)|. \quad (10)$$

One bit of uncertainty is equivalent to the total uncertainty regarding the truth or falsity of one atomic proposition. When the Hartley function H is applied to subsets of a given universal set \mathbf{X} , it has the form $H : \mathcal{P}(\mathbf{X}) \rightarrow \mathbb{R}^+$, where \mathbb{R}^+ denotes the set of nonnegative real numbers. In this case, its range is:

$$0 \leq H(\mathbf{V}(s)) \leq \log_2 |\mathbf{X}|. \quad (11)$$

Dubois and Prade [11] proposed non-specificity in evidence theory based on Hartley function:

$$N(p) = \sum_{\mathbf{V}(s) \in \mathcal{F}} p(\mathbf{V}(s)) \log_2 |\mathbf{V}(s)|. \quad (12)$$

Function N is a weighted measure by using Hartley function for all focal elements. The weights are values of the basic probability assignments. For each focal element $\{\mathbf{V}(s), p(\mathbf{V}(s))\} \in \mathcal{B}$ indicates the degree of evidence focusing on $\mathbf{V}(s)$, while $\log_2 |\mathbf{V}(s)|$ indicates the lack of specificity of this evidential claim. The larger the value of $p(\mathbf{V}(s))$, the stronger the evidence (in other words, the less unlinkability); the larger the set $\mathbf{V}(s)$ (i.e., the longer the packet forwarding

path), the less specific the evidence (i.e., the more unlinkability).

We denote the ordered set $\mathbf{V}(s_n)$ where the sequence $s_n = \{x_1, x_2, \dots, x_n\}$. Let $\mathbf{V}(s_1) \subset \mathbf{V}(s_2) \subset \dots \subset \mathbf{V}(s_n)$ be a complete sequence of nested subsets where $\mathbf{V}(s_k) = \{x_1, x_2, \dots, x_k\}$ for $k \in \mathbb{N}_n$. The focal elements in \mathcal{F} are nested and they are linearly ordered by the subset relation. Let $p_k = p(\mathbf{V}(s_k))$ and $\gamma_k = \gamma(x_k)$ for all $k \in \mathbb{N}_n$, where $\gamma : X \rightarrow [0, 1]$ is the possibility distribution [6]. Then, the n -tuples

$$\mathbf{P} = \langle p_1, p_2, \dots, p_n \rangle \quad (13)$$

fully characterize the basic probability assignments. Correspondingly, we refer the following probability assignments as possibility distribution:

$$\gamma = \langle \gamma_1, \gamma_2, \dots, \gamma_n \rangle. \quad (14)$$

The nested structure implies that $\gamma_j \geq \gamma_{j+1}$ for all $j \in \mathbb{N}_n$. That is, possibility distribution are in this formulation always ordered and $\gamma_1 \leq 1$ in each possibility distribution. It is easy to show that

$$\gamma_j = \sum_{k=j}^n p_k. \quad (15)$$

$$p_j = \gamma_j - \gamma_{j+1} \quad (16)$$

for all $j \in \mathbb{N}_n$, where $\gamma_{n+1} = 0$. From (13) to (16), we can derive the following assertions:

- $\gamma_1 = \gamma(x_1) = 1$ means that the packet originator x_1 is always on the evaluated path. This is intuitive and always true.
- γ_j is the summation of probabilities that are derived from the evidence for the path beyond node x_j .
- $\gamma_j \geq \gamma_{j+1}$ means that the longer path will cause less certainty of an evaluated path.

The possibility measure γ is another form of (or a transform of) the belief measure Bel in (5). In the possibility measure, sets are arranged in a nested manner $\mathbf{V}(s_1) \subset \mathbf{V}(s_2) \subset \dots \subset \mathbf{V}(s_n)$; in the belief measure, the evidence $\forall \mathbf{U} \subseteq \mathbf{V}$ contributes to the measure $Bel(\mathbf{V})$, however \mathbf{U}_s are not necessarily to be arranged in a nested sequence. The representations between the possibility measure and belief measure are also different: if we call the representation of belief measure a forward notation, the representation of possibility measure is a backward notation. For example, γ_1 represents the measure for one element (the source) in the path, but the probabilities contributed to γ_1 are from all the nested path segments (or subsets). The last difference between the belief measure and possibility measure is that $Bel(\mathbf{V}) \leq 1$ while $\gamma_1 = 1$. The above described properties shows that the possibility measure is very suitable for evaluating a communication path in MANET.

Using (15) and (16), we can rewrite (12) as follows:

$$N(\gamma) = \sum_{j=1}^n (\gamma(x_j) - \gamma(x_{j+1})) \log_2 |\mathbf{V}(s_j)|. \quad (17)$$

Equation (17) is the unlinkability measure for the path specified in $\mathbf{V}(s_n) = \{x_1, \dots, x_n\}$.

3.3 Unlinkability Measure for an MANET Communication System

We explore the statistical unlinkability evaluation for a communication system using Shannon-like information evaluations. Using evidence theory, the Shannon information theory based solutions [12], [13] can be presented as

$$S(p) = - \sum_{\{x\}} p(\{x\}) \log_2 Bel(\{x\}), \quad (18)$$

where the set $\{x\}$ is singleton and $Bel(\{x\}) = p(\{x\})$ is the belief measure for each mobile node x .

To measure the MANET system uncertainty, we propose the following MANET system unlinkability measure as follows:

$$E(\beta) = - \sum_{\mathbf{V}_{\langle o,d \rangle} \in \mathcal{F}} \beta(\mathbf{V}_{\langle o,d \rangle}) \log_2 \beta(\mathbf{V}_{\langle o,d \rangle}) \quad (19)$$

$E(\beta)$ is a generalized function to measure unlinkability of a given communication system in the number of “bits”. $E(\beta)$ can be reduced to Shannon measure when all evaluated sets are singletons. However, it is more general than Shannon measure in that it can be used to measure unlinkability for any given communication scenarios with derived body of evidence $\mathcal{B} = \{\mathcal{F}, \beta\}$. It must be noted that the key component in our proposed unlinkability measuring approach is to collect the evidence and derive the BP probability assignment function $\beta(\cdot)$ for each evaluated communication relation (e.g., represented by an inclusive set). To compute the system unlinkability, we need to take the following steps to compute $E(\beta)$:

1. Compute the traffic matrices construction based on the Definition 4;
2. Compute the communication-relation matrix based on the Definition 5;
3. Compute the BP probability assignments based on the Definition 6;
4. Finally, compute (19).

4 Evaluation Model

In this section, we propose two schemes to evaluate the performance of an unlinkability measure.

4.1 Amplification Ratio

In Definition 5, within the time period $t_K = \sum_{k=1}^K \Delta t_k$, we summarize the properties of the communication relation matrix \mathbf{R}^{t_K} as follows:

- (a) \mathbf{R}^{t_K} is a $N \times N$ square matrix, and each element $r_{\langle i,j \rangle}^{t_K} > 0$ represents the amount of traffic from node i to node j during the time period t_K . The diagonal of \mathbf{R}^{t_K} is the amount of traffic sent by node i , where $\forall i, j \in \mathbf{X}$ and $r_{\langle i,j \rangle}^{t_K} \leq r_{\langle i,i \rangle}^{t_K}$.
- (b) $r_{\langle i,j \rangle}^{t_K}$ is the maximum accumulative traffic from node i to node j via all possible paths. We have the following observations:

- (i) $\mathbf{R}_{i,*}^{t_K} = \sum_{j \neq i} r_{\langle i,j \rangle}^{t_K}$ is the maximum traffic to all potential receivers from node i .
- (ii) $\mathbf{R}_{*,j}^{t_K} = \sum_{i \neq j} r_{\langle i,j \rangle}^{t_K}$ is the maximum traffic to one receiver j from all possible sources.
- (iii) $\mathbf{R}_{*,*}^{t_K} = \sum_{\forall i} \sum_{j \neq i} r_{\langle i,j \rangle}^{t_K}$ is the maximum traffic among all possible pairs.
- (iv) $\mathbf{R}_{i,i}^{t_K} = \sum_{\forall i} r_{\langle i,i \rangle}^{t_K}$ is the actual traffic transmitted in the system.

The actual network traffic $\mathbf{W}|_{1 \times K}$ is amplified in the procedure to derive the communication relation matrix \mathbf{R}^{t_K} . The amplification is due to the uncertainty in deriving the end-to-end communication relations. In other words, the amount of amplified information is the amount of introduced *noise*. We define the ratio of the traffic amplification as a metric to evaluate the amount of noise information to confuse the adversaries. We define the following three amplification ratios:

$$\mathcal{A}_{i,*}^{t_K} = \mathbf{R}_{i,*}^{t_K} / r_{\langle i,i \rangle}^{t_K}, \quad (20)$$

$$\mathcal{A}_{*,j}^{t_K} = \mathbf{R}_{*,j}^{t_K} / \alpha_{\langle j,j \rangle}^{t_K}, \quad (21)$$

$$\mathcal{A}_{*,*}^{t_K} = \mathbf{R}_{*,*}^{t_K} / \mathbf{R}_{i,i}^{t_K}. \quad (22)$$

(20) is the sender amplification ratio from one source to all possible destinations; (21) is the receiver amplification ratio from all possible sources to one destination; (22) is the system amplification ratio, i.e., the total deduced traffic to the total actual traffic. For single transceiver wireless communications, an actual receiver remains silent during the transmission. Here, we use $\alpha_{\langle j,j \rangle}^{t_K}$ to represent the actual traffic accepted by node j which is different from the definition of $r_{\langle i,i \rangle}^{t_K}$. It is difficult to derive the exact actual traffic $\alpha_{\langle j,j \rangle}^{t_K}$ in (21) accepted by node j . Thus, our following discussions focus on the amplification ratios defined in (20) and (22).

4.2 Measuring Diversity

Shannon information theory based unlinkability measure is shown in (18). As shown in Definition 2, perfect unlinkability is achieved when the communication system is in an indistinguishable state. That is every event X (among all N possible events) has the same occurrence frequency $1/N$. It is useful to compare the actual unlinkability of a communication relation to its maximum achievable unlinkability. We propose to use Kullback-Leibler divergence [7] (denoted by KL, a.k.a., information divergence, or information gain, or relative entropy) for this purpose:

$$KL(\mathbf{BP}||\mathbf{Q}) = \sum_X p(x) \log_2 \frac{bp(x)}{q(x)}, \quad (23)$$

where $\mathbf{BP} = (\beta(\mathbf{V}_{\langle o,d \rangle}))_{N \times N}$ and $bp(x) = \beta(\mathbf{V}_{\langle o,d \rangle})$ w.r.t an arbitrary probability distribution $\mathbf{Q} = \{q(x)|x = |\mathbf{X}|\}$ and $|\mathbf{X}| = N$. Using the property of perfect unlinkability, we require $\forall x, q(x) = 1/N$. KL (a.k.a., information divergence, or information gain, or relative entropy) is a natural distance measure from a *true* probability distribution \mathbf{BP} to

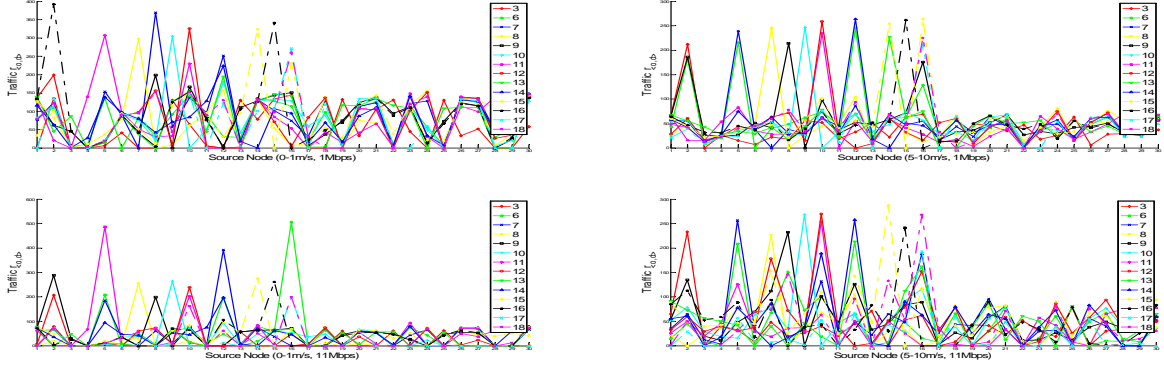


Figure 2: Traffic volume: from a source node to each receiver.

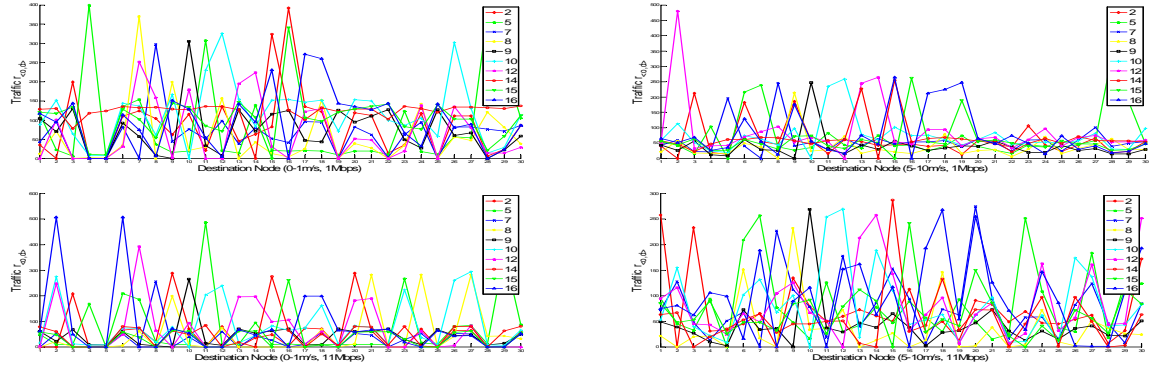


Figure 3: Traffic volume: from each source node to one destination node.

an arbitrary probability distribution \mathbf{Q} . Typically \mathbf{BP} represents data, observations, or a precise calculated probability distribution. The measure \mathbf{Q} typically represents a theory, a model, a description or an approximation of \mathbf{BP} . In our case, it is the maximum unlikability. KL measure represents the *distance* from an actual unlikability measure to the theoretical maximum unlikability that the system can provide. We can conclude that smaller the distance is, more the unlikability provides.

5 Experiments

In this section, we use experiments to illustrate the application of the proposed unlikability measure to an IEEE 802.11b based MANET, as well as the use of the evaluation model. We use Qualnet [14] to conduct the simulation study. We perform a comparative study which includes 4 scenarios indexed by moving speed ranges and node transmission rates as shown in the following table.

Tran. Rate	Moving Speed	
	0m~1m/s	5m~10m/s
1Mbps/s	radius=483.741m, Rev. sensitivity=93dBm	
11Mbps/s	radius=283.554m, Rev. sensitivity=83dBm	

The simulation time period is 200 seconds. Four scenarios follow the same simulation construction given as follows: (a)

30 wireless nodes are randomly deployed in an $800 \times 800m^2$ simulation area; (b) 15 end-to-end pairs (i.e., 2-3, 5-6, 5-7, 7-8, 8-9, 8-10, 9-10, 10-11, 10-12, 12-13, 12-14, 14-15, 15-16, 16-17, 16-18) are randomly chosen to transmit CBR traffic, where 10 nodes originate data frames (i.e., 2, 5, 7, 8, 9, 10, 12, 14, 15, 16) to 14 sinks (i.e., 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18). All other nodes are packet forwarding nodes. (c) To derive the traffic-communication relation matrix $\mathbf{R}^{t\kappa}$ using the transformation algorithms presented in Section 3.1, we set the timing threshold $\mathcal{T} = 4s$ to filter out the packets “travel” too long and set the hop count threshold $\mathcal{H} = 5$ to prevent the packets “travel” to many hops. In this way, we can prevent unreasonable “long” paths.

5.1 End-to-end Unlikability Measure

The communication-relation matrix (presented in the Definition 5) is used to derive the BP measure. $r_{(o,d)}$ represents the accumulated (i.e., number of packets) evidence supporting the end-to-end communication relation $o \rightarrow d$. Higher value of $r_{(o,d)}$ represents better linkability from o to d . In (4), the row vector represents the traffic from one sender to every receiver (shown in Figure 2). Each curve represents the total amount of end-to-end traffic originated from a source node (indexed by x -axis) to each sink (indexed by the legend). The column vector in (4) represents the traffic received from multiple senders to one receiver (shown in Figure 3). Each curve

represents the total amount of end-to-end traffic received by a sink (indexed by x-axis) from each source node (indexed by the legend). In these figures, we notice that the peaks can be classified as high peaks and lower peaks. Most of high peaks represent real packet src/dst pairs (see the high peak values that map to the real src/dst pair list in the beginning of this section) and the low peaks usually represent the packet forwarding relations.

In low speed scenarios, the distances among mobile nodes are unchanged for a relatively longer time and the path between each pair of end nodes is also stable. We find that 1Mbps case exhibits better unlinkability than 11Mbps case, where the high peaks are easily distinguishable from low peaks. This is because the low transmission rate has long transmission distance due to wireless signal properties. As a result, the number of potential receivers covered by a 1Mbps sender is larger than an 11Mbps sender. Then the transformation function f presented in Section 3.1 will generate more deduced traffic due to the transformation function g .

In high speed scenarios, the path changes are frequent. Thus a forwarding node has a better chance to serve as an intermediate node for multiple communication sessions. In both Figure 2 and Figure 3, the 11Mbps case causes more forwarding traffic (i.e., low peaks are a little higher) than the 1Mbps case. This is because the higher transmission rate, the shorter transmission distance, and thus the more path changes will occur. In a summary, comparing the low speed and high speed cases, we find that high speed will increase the unlinkability under traffic analysis attacks.

5.2 Unlinkability Measure of a Communication Path

Using (17), we derive the path unlinkability measure for different paths shown in Figure 4. The higher non-specificity measure means better unlinkability. Thus, we conclude that longer path is preferred than shorter path to achieve better unlinkability.

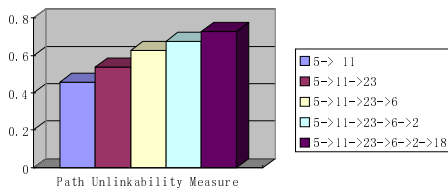


Figure 4: Unlinkability Measure of One-path Communications.

5.3 Unlinkability Measure of the MANET System

Using (19), we draw the BP system measure in Figure 5. The higher BP system unlinkability measure means better unlinkability of the MANET system. From the figure, we conclude that (a) the BP system unlinkability measure confirms our previous finding “slower transmission rate will increase unlinkability”, and (b) high speed will also improve the MANET system’s unlinkability.

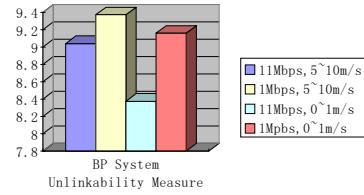


Figure 5: BP Unlinkability Measure of an MANET Communication System.

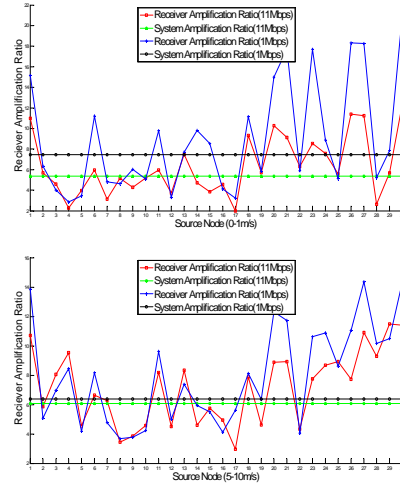


Figure 6: Amplification Ratio.

5.4 Amplification Ratio and KL Measure

In Figure 6, amplification ratios for each source node and the overall system have been presented. The system amplification ratio is the average of the receiver amplification ratios. High amplification ratio means better unlinkability. We notice that both receiver and system amplification ratios of high transmission rate are lower than low transmission rate, which means less unlinkability. This result confirms the previously studied unlinkability measure. In addition, high speed exhibits better unlinkability than low speed.

Figure 7 is derived from the probability assignments defined in (23). The smaller KL value means the closer to the maximum unlinkability presented in the Definition 2. When the KL measure equals to zero, the ultimate unlinkability is achieved. The peak values of each KL curve represent that the corresponding source nodes have worse unlinkability than other source nodes. The high speed also reduces the KL measure value which again confirms our previously conducted simulation studies.

5.5 Summary

We have shown how to use various unlinkability measures to assess the unlinkability from different angles of an MANET. In this research, the main findings are: (a) transmission rate is one of main factors that affect the unlinkability, i.e., lower transmission rate exhibits better unlinkability, (b) mobility is another main factor that affects the unlinkability, i.e., high

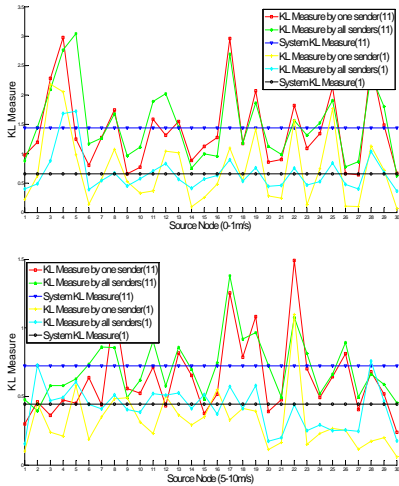


Figure 7: KL Measure.

speed increases the unlinkability, and (c) longer path is preferred to improve unlinkability. From the above findings, we note that increasing unlinkability usually means reducing the system's performance in terms of capacity.

6 Conclusion

In this paper, we have the following observations which can guide us in future unlinkability research:

(i) Evidence theory is a generalized information theory that can be reduced to Shannon information theory under certain circumstances. Compared to Shannon information theory, evidence theory relies on a well founded evidence collection mechanism. In communication systems, different scenarios (such as protocols, hardware devices, etc.) need different appropriate measuring and collecting methods.

(ii) In addition to the number of captured data frames and timing, evidence has broader meanings, such as mobility. We need to better exploit the uses of our model on different types of evidence.

(iii) Validated by our simulation study, we conclude the following findings: (a) lower transmission rate exhibits better unlinkability, (b) high speed increases the unlinkability, and (c) longer path is preferred to improve unlinkability. Thus efficient anonymous MANET communication protocols must be devised to consider both unlinkability and QoS requirements.

REFERENCES

- [1] C. Díaz, "Anonymity and privacy in electronic services," Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, December 2005.
- [2] S. Steinbrecher and S. Kopsell, "Modelling Unlinkability," *Privacy Enhancing Technologies Workshop (PET) 2003*, 2003.
- [3] G. Tóth, Z. Hornák, and F. Vajda, "Measuring Anonymity Revisited," in *Proceedings of the Ninth Nordic Workshop on Secure IT Systems*, S. Liimatainen

and T. Virtanen, Eds., Espoo, Finland, November 2004, pp. 85–90.

- [4] D. Huang, "Traffic Analysis-based Unlinkability Measure for IEEE 802.11b-based Communication Systems," in *Proceedings of ACM Workshop on Wireless Security (WiSe)*, 2006, pp. 65–74.
- [5] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [6] G. J. Klir and M. J. Wierman, *Uncertainty-Based Information*. Physica-Verlag, A Springer-Verlag Company, 1998.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [8] A. Pfitzmann and M. Hansen, "Anonymity, unlinkability, unobservability, pseudonymity, and identity management - a consolidated proposal for terminology," *Working draft, available at http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.28.doc*, September 2006.
- [9] R. V. L. Hartley, "Transmission of Information," *The Bell System*, vol. 7, no. 3, pp. 535–563, 1928.
- [10] Y. Zhang, W. Liu, W. Lou, and Y. Fang, "MASK: Anonymous On-demand Routing in Mobile Ad Hoc Networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2376–2385, 2006.
- [11] D. Dubois and H. Prade, "A Note On Measures of Specificity for Fuzzy Sets," *International Journal of General Systems*, vol. 10, no. 4, pp. 279–283, 1985.
- [12] C. Díaz, S. Seys, J. Claessens, and B. Preneel, "Towards Measuring Anonymity," in *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002), Lecture Notes in Computer Science*, vol. 2482. Springer, 2002, pp. 54–68.
- [13] A. Serjantov and G. Danezis, "Towards an Information Theoretic Metric for Anonymity," in *Proceedings of Privacy Enhancing Technologies*. Springer, 2002, pp. 41–53.
- [14] QualNet Communications, "QualNet Simulator, <http://www.qualnetcomm.com/>."