

# A location privacy metric for V2X communication systems

Zhendong Ma, Frank Kargl, and Michael Weber  
Institute of Media Informatics, Ulm University, Germany  
{zhendong.ma|frank.kargl|michael.weber}@uni-ulm.de

**Abstract**—The emerging vehicle-to-vehicle/vehicle-to-infrastructure (V2X) communication systems enable a new way of collaboration among the vehicles, the operators of transportation systems, and the service providers. However, many functionalities of V2X systems rely on detailed location information. This raises concerns on location privacy of the users of such systems. Although privacy protection mechanisms have been developed, existing privacy metrics are inappropriate or insufficient to reflect the true underlying privacy values in such systems. Without a proper metric, preserving location privacy in V2X communication systems becomes difficult due to the lack of a benchmark to evaluate any given protection mechanisms. In this paper, we develop a quantitative metric to measure and quantify location privacy in V2X systems. To do so, we introduce the concept of snapshots, which capture the information related to a user in a given space and time. Then the level of location privacy is quantified as the uncertainty of the information related to that user. Our analyses show that the metric provides the users, the system designers, and other stakeholders a valuable tool to evaluate the risk and measure the level of location privacy in V2X communication systems.

## I. INTRODUCTION

The emerging vehicle-to-vehicle/vehicle-to-infrastructure (V2X) communications creates enormous interests and efforts in research and prototype development, due to the prospect of a safer, more convenient and efficient transportation system in the near future. If deployed, the Dedicated Short Range Communication (DSRC) based V2X systems will become the biggest realization of mobile ad hoc networks (MANET). Example applications of V2X systems include intersection collision warning, traffic monitoring through probe vehicle data, and location based services etc.

However, many envisioned V2X applications rely on detailed and continuous vehicle location information. Vehicles are highly personal devices, the sending and dissemination of personal location information has the potentials to infringe a user's privacy, especially its location privacy. The privacy issue in V2X systems has been identified and mechanisms have been proposed to preserve user privacy, e.g., in [1], [2].

To evaluate the performance of any proposed privacy enhancing technologies (PET) in V2X communications, we need a privacy metric which can measure and reflect the achieved privacy protection. A privacy metric might also be wanted by privacy advocacy groups, legislators, and consumers to

evaluate whether the V2X communication systems comply with the current social and legal mandates on personal privacy.

However, existing privacy metrics are inappropriate and insufficient to reflect the underlying value. More specifically, they are either only applicable for systems with very different characteristics, or unable to capture information on the whole. For example, metrics on user identity do not take considerations of user movements, and metrics on user movements leave out the information on user identity. Moreover, most of the existing metrics stay at the theoretic level, which provides only small practical values.

In this paper, we develop a location privacy metric for V2X communication systems. The metric measures and quantifies the level of location privacy of each user in the system. The key idea is to measure the level of location privacy as the *linkability* of location information to the individuals who generate it. Doing so is partially inspired by the Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data [3], which states that *personal data* is any information relating to an identified or identifiable natural person. As an interpretation, we regard an individual's location information as 'personal information', thus it cannot be used to identify or link to the individual without his or her consent.

Our main contribution is the development of a location privacy metric and the methodology to enable the measurement. We introduce the concept of *snapshots*, which capture the information related to user location privacy and encapsulate it into a measurement model. Taking an information theoretic approach, the privacy level is quantified as the uncertainty of relating location information to a particular individual. Our analyses show that the metric provides a valuable tool to evaluate the location privacy risk and measure the level of location privacy in V2X communication systems.

In the following, Sec. II discusses the existing privacy metrics and their limitations. Sec. III analyzes the concept and related issues of location privacy. Sec. IV introduces the metric. Sec. V applies and evaluates the metric by means of case studies. Sec. VI concludes the paper and points out the direction of our future work.

## II. EXISTING PRIVACY METRICS

The level of privacy can be expressed in terms of anonymity. Anonymity describes a user's anonymous state regarding a

specific action. Anonymity can also be expressed as unlinkability between a user and a specific action (e.g., sending a message). The definitions of anonymity and unlinkability are given in [4] and further refined in [5]. In this paper, we also use 'linkability' to refer to the opposite of unlinkability, and 'linking' or 'linkage' for the action or the status of linkability.

User anonymity is often measured by the size of anonymity set in anonymous communication systems. The authors of [6], [7] find out that the size of the anonymity set does not reflect different probabilities of the members in the set, and propose to use entropy as the metric for anonymity. In wireless communications, a user usually does not have a fixed anonymity set due to its dynamic changes of the neighboring status with other nodes. Beresford et. al [8] use mix zones to model user movements and quantify the user anonymity as the entropy of an adversary's probability assignments on the user movements through the mix zones. Authors in [9] and [10] also use the entropy of the mix zones to evaluate location privacy achieved by the vehicles in vehicular ad hoc networks (VANET). Mix zones cover only parts of the whole geographic area where V2X systems will be deployed. Therefore, location privacy measurements based on the entropy of mix zones cannot capture all information of V2X systems.

Tracking is another common approach to measure the level of privacy. Gruteser et al. [11] propose to use tracking algorithms to characterize the level of location privacy. Sampigethaya et. al [12] use maximum tracking time to evaluate the location privacy of vehicles in VANET. Hoh et. al [13] use the mean time to confusion to measure the privacy level of vehicles sending GPS traces in a traffic monitoring system. Their metric is based on an entropy measure of the adversary's tracking uncertainty. We argue that without linking the location traces to an individual, tracking alone only partially achieves the goal of any attacks on location privacy. Thus it is insufficient as a metric for location privacy.

Fischer et. al [14] find out that entropy measures are insufficient for measuring unlinkability when the sender-message relations are modeled by set partitions of observed messages. They propose to use a measure taking considerations of both the outer and inner structures of the set partitions. Although the approach is very promising, it would be interesting to see how it can be scaled to apply to real systems.

### III. LOCATION PRIVACY IN V2X COMMUNICATIONS

Personal information can be modeled as a person's context information. Context is any information that can be used to characterize the situation of a person, and the primary contexts are *identity, location, activity and time* [15]. Obviously, context on location, activity and time only becomes personal and privacy relevant when identity information is associated. Given the fact that a person's activity can often be derived from location and time, if an unauthorized party can link location and time information to a person's identity, the context is complete and the individual's privacy is at stake.

In V2X communications, location information is either explicitly or implicitly given in the outgoing messages. We

can generalize them as location samples, which contain information on *identifier, location and time*. The information on identifier, although not necessarily to be a person's identity, can help to identify an individual or the message relations. Location and time are either explicitly given in a message or implicitly derived from the place and time of the messages recorded. Each time a user sends a message, it leaves a 'digital footprint' in the system. An adversary on location privacy tries to follow the vehicle movements based on location samples. The adversary can exploit such information to identify and profile a user, or even to infer the activities or any sensitive information about the user. Depending on whether the adversary is outside or inside the system, the methods of obtaining location samples range from eavesdropping communications to directly accessing the data stored in the system. Since adversaries vary in capacities, the obtained location information vary in qualities. Despite the variations, the location information can be categorized into three types: 1) *a single location*; 2) *a track*, a sequence of locations from the same vehicle, which reveals a vehicle's movements in space and time; and 3) *a trip*, a complete journey of a vehicle from the origin to the destination. A trip is an ensemble of tracks.

Given a single location, it is very difficult to link it to a specific user unless the identity information is included in the message. An adversary can obtain tracks by linking a sequence of messages with same identifiers, or by one of the target tracking methods [16]. A track only provides partial information on a vehicle's movements. However, if the adversary is able to connect all tracks of the same vehicle, it turns the tracks into a trip. A trip contains information on the time, the origin (O) and the destination (D) of a journey. Often it can be used to infer an individual's identity and activities. The empirical study [17] shows that given the trip information, it is possible to heuristically infer the home address and the identity of the driver of a vehicle.

For privacy concerns, V2X communication systems are very likely to employ some pseudonym systems like [1], [18] to protect a user's identity. Consequently, an adversary has to rely on the trip information to learn the identity and activities of the driver. Since trips contain information on locations and tracks, as well as information to infer further information, they are our main focus in the metric.

### IV. LOCATION PRIVACY METRIC

A metric is a system or standard of measurement. We are interested in a quantitative measure reflecting location privacy of the users involved in V2X communications. Our analysis shows that in the context of V2X communications, the properties of location privacy consist of individuals and their trip information. Therefore, we can use an adversary's ability to link vehicle trips to specific individuals to reflect the level of location privacy of the individuals.

#### A. Methodology

A V2X communication system is a dynamic system and continuous in space and time. To allow us to take a sensible

measurement of the system, we need to take a discrete sample from the system and base our measurement on a relatively static and confined version. Thus we make three assumptions:

- 1) The location information considered in the metric is assumed to be within an arbitrarily defined area.
- 2) The location information considered in the metric is assumed to be within an arbitrarily defined time period.
- 3) We further assume that the adversary is able to identify a location as the origin or destination of a trip.

Combining these three assumptions, we derive that there will be only complete trips and the number of origins and destinations are equal. The first two assumptions enable us to virtually take a *snapshot* of the system. The snapshot captures the vehicle movements and their relations to the drivers in a given area and time period. As a first step, we only consider a *single* snapshot in this paper. In our future work, we plan to include consecutive snapshots in the metric. So the above assumptions will be relaxed to allow multiple snapshots.

To quantitatively measure user location privacy, we take the following steps. In the first step, we model the information contained in the snapshot in a *measurement model*. The measurement model abstracts the location information into three basic components: the linkage of an individual to an origin of a trip, the linkage of an origin to a destination, and the linkage of a destination to an individual. The adversary's knowledge of the system is expressed as probability assignments on each of the linkages. In the second step, the probability distributions in the measurement model are extracted to yield quantitative measurements.

### B. The measurement model

Modeling the information contained in the snapshot is to represent the information in an abstract and mathematical form to facilitate calculation in the next step. We observe that the information in the snapshot contains the information on individuals, O/D pairs, and their interrelations. We also observe that for an individual to 'make a trip', he or she must start a trip at an origin and ends the trip at a destination. This also implies that the individual at the origin and the destination should be the same person.

Based on the observations, we model the information as a weighted directed graph  $G = (V, E, p)$ . There are three disjoint sets of vertices in the digraph, i.e.,  $I \subseteq V$ ,  $O \subseteq V$ , and  $D \subseteq V$  with  $I \cup O \cup D = V$ .  $I$  is the set of all individuals,  $|I| = n$ .  $O$  is the set of all origins and  $D$  is the set of all destinations of the trips,  $|O| = |D| = m$ . The edge set  $E$  is defined as  $E := E_1 \cup E_2 \cup E_3$  with  $E_1 := \{e_{io} | i \in I, o \in O\}$ ,  $E_2 := \{e_{od} | o \in O, d \in D\}$  and  $E_3 := \{e_{di} | d \in D, i \in I\}$ . As  $E_1, E_2, E_3$  are disjoint,  $G$  is a tripartite graph. Each edge  $e_{jk} \in E$  is weighted by a probability function  $p: E \mapsto [0, 1]$ .

$G$  has several notable properties. First,  $G$  contains all aforementioned information in the snapshot. Since tracking is not the focus of the paper, we assume that there is a publicly known tracking algorithm and treat vehicle tracking as a black box, i.e. we assume that  $p(o_j, d_k)$  is known. Second, vertices in  $G$  are connected with directed edges. If we follow

the directed edges from a vertex  $i_s$ , the path will pass the vertices  $\{i_s, o_j, d_k, i_s\}$ . The semantics of the cycle is  $i_s$ 's possibility having made a trip from  $o_j$  to  $d_k$ . Third, the probability distributions on the edges model an adversary's knowledge of the users and their movements in the system<sup>1</sup>. In addition, we define that the sum of the probabilities on outgoing edges from a vertex  $o \in O$  or  $d \in D$  to be 1,  $\sum_{k=1}^m p(o_j, d_k) = 1$ ,  $\sum_{k=1}^n p(d_j, i_k) = 1$ , while letting the sum of probabilities from the vertex  $i \in I$  to be equal or smaller than 1,  $\sum_{k=1}^m p(i_j, o_k) \leq 1$ . By the latter definition, we model an individual not making any trips. For example,  $\sum_{k=1}^m p(i_1, o_k) = 0.9$  means that  $i_1$  has 0.9 probability to make trips and 0.1 probability to 'stay at home'.

For the ease of calculations, we also represent  $G$  by three adjacency matrices, **IO**, **OD**, and **DI**. Each entry  $a_{jk}$  in the matrices indicates that there is an edge from vertex  $v_j$  to vertex  $v_k$ . The value of the entry is the weight on the edge,  $a_{jk} = p(v_j, v_k)$ . Furthermore, each row in the matrices is a vector of the probability distribution on all outgoing edges from the same vertex. The sum of each row in **IO** is equal or smaller than 1, and the sum of each row in **OD** and **DI** equals 1.

### C. Calculation

To extract the probability distributions and quantify the information in the measurement model, we use information entropy developed by Shannon [19]. Entropy is a quantitative measure of information content and uncertainty over a probability distribution. Entropy has been widely accepted as an appropriate measure in the privacy research community [6]–[8]. However, the main challenge here is to apply entropy to the measurement model.

By definition, for a probability distribution with values  $p_1, \dots, p_n$ , the entropy is

$$H = - \sum p_i \log p_i \quad (1)$$

where  $p_i$  is the  $i^{\text{th}}$  element of the probability distribution.  $H$  is then a measure of information content and uncertainty associated with the probability distribution. The logarithm in the formula is usually taken to base 2 to have a unit of *bit*, indicating how many bits are needed to represent a piece of information. A higher entropy means more uncertainty and hence a higher level of privacy. Entropy reaches its maximum value if all the probabilities in the distribution are equal.

Shannon uses entropy as a quantitative measure of the information produced by a discrete information source. When applying entropy to our calculation, the source is the information captured in the measurement model accessible to the adversary. We are interested in the information on the relations between the individuals and the trips, i.e., the information on who moves from where to where. The information is expressed as the probabilities of a particular individual within the system to make one of the trips, as well as to not make any trips.

We are interested in the entropy (the uncertainty) related to an individual and the  $m$  O/D pairs (which leads to  $m^2$  possible

<sup>1</sup>How the adversary obtains the probabilities will be discussed in Sec. V.

trips) in the system. If we 'unfold' all the cycles related to a particular individual in  $G$  (cf. Sec. IV-B), we obtain a flower-like structure shown in Fig. 1(a). The stigma, or the center of the flower is the individual, e.g.,  $i_1$ . The petals run clock-wise around the stigma, denoting  $i_1$  making one of the  $m^2$  possible trips, with the last petal representing  $i_1$  does not make a trip. We denote this complementary probability  $p^c$ . If we assume that the measurements reflect separate observations, i.e., the probabilities describe independent events, the probability of an individual making a specific trip is the product of the probabilities on all edges of the petal representing that trip. We can further simplify the flower structure to the wheel-like structure in Fig. 1(b). The hub in the center represents an individual, e.g.,  $i_1$ . Each radiating spoke from the hub represents the probability of  $i_1$  making a specific trip.

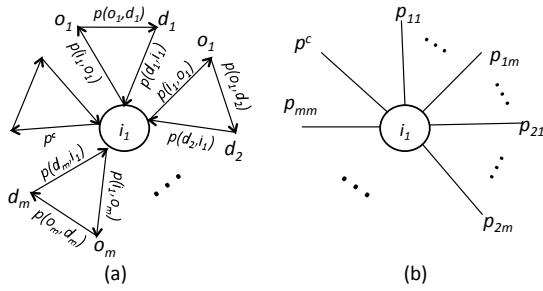


Fig. 1. Examples of visualizing the probability distribution related to an individual as (a) a flower, and (b) a 'hub and spokes'

We take the non-zero probabilities and normalize them to calculate the entropy, because  $p_i = 0$  means there is no uncertainty and the sum of the probability distribution should equal 1. Based on Formula (1) and using the notation specified in the measurement model, we calculate the entropy for a specific individual as

$$H(i_s) = -\left(\sum_{j=1}^m \sum_{k=1}^m \hat{p}_{jk} \log(\hat{p}_{jk}) + \hat{p}^c \log(\hat{p}^c)\right) \quad (2)$$

where  $\hat{p}_{jk}$  is the normalized probability of  $i_s$  making a trip from  $o_j$  to  $d_k$  and  $\hat{p}^c$  is the normalized probability of  $i_s$  not making any trips. The values of  $\hat{p}_{jk}$  and  $\hat{p}^c$  are given as

$$\hat{p}_{jk} = \frac{p(i_s, o_j)p(o_j, d_k)p(d_k, i_s)}{\sum_{j=1}^m \sum_{k=1}^m p(i_s, o_j)p(o_j, d_k)p(d_k, i_s) + \hat{p}^c} \quad (3)$$

$$\hat{p}^c = 1 - \sum_{j=1}^m p(i_s, o_j) \quad (4)$$

To evaluate the location privacy of an individual, it is also useful to find the maximum entropy possible for an individual in the system, i.e., the upper bound. The maximum entropy for an individual is reached if all of the participants in the system are equiprobable to make any trips and all trips are also equiprobable. In a system with measurements of  $m$  O/D

pairs, the maximum entropy of an individual  $i_s$  is

$$\text{Max}H(i_s) = -\log\left(\frac{1}{m^2 + 1}\right) \quad (5)$$

where 1 in the denominator accounts for the individual not making any trips. Interestingly, the maximum entropy for an individual depends only on the number of possible trips, not the number of participants in the system.

Given the entropy upper bound, the level of location privacy of an individual can also be expressed as the ratio of the current entropy to the maximum. Therefore, we have

$$H\% = \frac{H(i_s)}{\text{Max}H(i_s)} 100\% \quad (6)$$

which uses % as the unit. We use  $H\%$  to express the ratio of an individual's privacy level to the maximum possible level. In other words, it gives a hint as how far an individual is from the theoretical privacy upper bound. Notice that  $H\%$  is different from a similar formula  $d = H(X)/H_M$  used in [7], which measures the degree of anonymity given an anonymity set.

## V. ANALYSIS

### A. Case study I

First, we use a simple example to illustrate how the metric works. Consider the scenario in Fig. 2. Three individuals  $i_1$ ,  $i_2$ , and  $i_3$  live on the same street, their homes are close to the locations  $h_1$ ,  $h_2$ , and  $h_3$ , from where three trips originating at almost the same time. The adversary's tracking result shows that the destinations of the trips are the university  $U$ , the hospital  $H$ , and the cafe  $C$ .

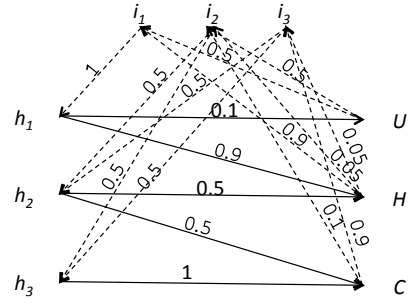


Fig. 2. A simple example

The probability assignments reflect the information the adversary obtained from observing the system. The adversary is sure that  $i_1$  starts from  $h_1$ , but thinks  $i_2$  and  $i_3$  are both probable to start from either  $h_2$  or  $h_3$ . The adversary also knows that both  $i_1$  and  $i_2$  work at the university, and  $i_1$  has visited the hospital quite often in the past. The adversary knows that  $i_2$  and  $i_3$  often go to the cafe. Since  $i_2$  is supposed to work at that time, the adversary assigns a higher probability to  $i_3$ . Besides, the adversary makes the probability assignments independently. For example, in the case of considering  $i_2$  making a trip from  $h_2$  to  $H$ , although the probability of  $i_2$  starting from  $h_2$  to  $H$  is  $0.5 * 0.5 = 0.25$ , it has no influence when the adversary assigning 0.05 as the probability of linking

$H$  back to  $i_2$ . By such assignments, we can model the situation in which later information influences the certainty of the whole trip. Another noteworthy assignment is the 1 on  $h_3$  to  $C$ . This happens when the adversary can track a complete trip, but cannot link the trip to a particular individual with certainty.

Using (2) – (6), we calculate the entropies and list the result in Tab. I. The result shows that  $i_1$  has the lowest entropy, hence the lowest privacy level. A close look at the example reveals that among all the possible trips,  $i_1$  can be linked to the trip from  $h_1$  to  $H$  with high certainty. As the uncertainty is low,  $i_1$ 's entropy becomes low. On the other hand,  $i_2$  has the highest entropy because the uncertainty is high to link  $i_2$  to the trips from  $h_2$  and  $h_3$ , as well as the destinations  $H$  and  $C$ . Although very simple, the example demonstrates that the metric is an effective tool to process various information and reflect the underlying privacy value.

TABLE I  
RESULT OF CASE STUDY I

$i_s$	$H(i_s)$	$H_{\%}$
$i_1$	0.32	9.6%
$i_2$	1.38	42%
$i_3$	1.03	31%

### B. Case study II

In the second case study, we analyze the role of tracking on location privacy. In this scenario, the adversary can track vehicles with high certainty, but has difficulties to link the vehicle movements to the individuals. It assigns higher probabilities to the individuals in the vicinity of the origins or destinations, and gradually decreases the probabilities as the individual's distances to the origins or destinations increase.

To simulate this scenario, we generate probabilities from a normal distribution for each row in matrices  $\mathbf{IO}$  and  $\mathbf{DI}$ , and probabilities from an exponential distribution for each row in matrix  $\mathbf{OD}$ . We simulate the scenario in MATLAB with 50 individuals and 20 O/D pairs. The probabilities are randomly generated from the probability distributions. Fig. 3 shows three example probability distributions from the three matrices. The first distribution is the probability of  $i_1$  starting at one of the 20 origins. Since the probabilities are taken from a normal distribution, they are quite evenly distributed around 0.05. The second distribution shows the probabilities of a vehicle from  $o_1$  ending at one of the 20 destinations. The probabilities are exponentially distributed, so several destinations have much higher values than the rest. The third distribution is also taken from a normal distribution. It shows the relations of the destination  $d_1$  to the 50 individuals.

Arbitrarily, we define an exponential distribution with  $\mu = 1$ , and a normal distribution with  $\mu = 0.5$  and  $\sigma = 0.1$ . The probabilities in each row of the matrices are randomly generated according to their distributions and normalized to 1. Then the three matrices are fed to the metric calculation. We repeat this process for 100 times, each time with three new randomly generated probability matrices. In the end, we

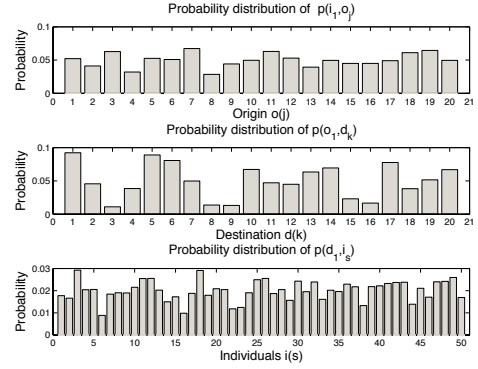


Fig. 3. Example of probability distributions in the three matrices.

obtain an average entropy over the 50 individuals over 100 simulations of 8.02 *bits*. As the maximum entropy for a system with 20 O/D pairs is 8.65 *bits*, we have a ratio  $H_{\%} = 92.7\%$ . When comparing to the values in case study I, the individuals in this scenario enjoy a very high level of location privacy. We will also interpret the result in the next section by comparing it to the result from case study III.

### C. Case study III

In the third case study, we try the opposite: high certainty on the linking of individuals to the origins and destinations, low certainty on tracking. To be able to compare with the results from case study II, we use the same setting of 50 individuals and 20 O/D pairs. We exchange the probability distributions. Specifically, we let matrices  $\mathbf{IO}$  and  $\mathbf{DI}$  have the probabilities from an exponential distribution, and  $\mathbf{OD}$  have the probabilities from a normal distribution.  $\mathbf{IO}$  and  $\mathbf{DI}$  simulate the situation that the adversary has more information on the individuals, such as where they live and what their daily schedules are, resulting in high probabilities on linking an individual to a few origins and linking a destination to a few individual. Due to poor tracking performance, the adversary has problems to link origins and destinations. This is simulated by probabilities from a normal distribution in  $\mathbf{OD}$ .

Using the same parameters for the exponential and normal distributions and the same process in case study II, we obtain an average entropy over the 50 individuals over 100 simulations of 7.48 *bits*, and  $H_{\%} = 86.5\%$ .

Fig. 4 compares the average entropy of the 50 individuals after each simulation run from both case studies. For all the simulation runs, entropies from case study II are higher than the ones from case study III, meaning that users in case study II have more location privacy than those in case study III. The entropy values fluctuate slightly, because the probabilities are re-generated at each simulation run. However, on the long run, they are quite stable around certain values. The result shows that the linkability of location information to particular individuals has more influences on the overall location privacy level than vehicle tracking. This means interestingly it will be more efficient to devise mechanisms to increase the unlinkability between location information and individuals.

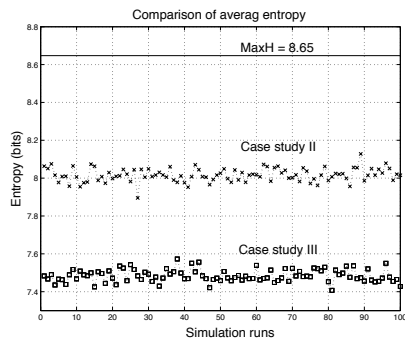


Fig. 4. Comparison of average entropy from case study II and III.

#### D. Discussion

One might ask how to assign probabilities so they reflect the true amount of information an adversary has on the system? In general, we can employ two approaches. Although having a list of all possible attacks on location privacy in V2X systems will be a NP-hard problem, in the first approach, we can derive the probabilities based on a) a set of already identified attacks, e.g., home identification and target tracking [20]; b) the information to be included in the communications of potential V2X applications; and c) publicly available data like land-use data and telephone directory. This can be a useful way to evaluate the conformance of a given V2X application to the privacy requirements. In the second approach, we can use probability mass functions to approximate the statistical data on population distributions and traffic statistics to have a large-scale analysis of V2X systems. In the above case studies, case study I employs the first approach and case study II & III employ the second approach.

The metric in this paper considers only complete trips. As a part of the location information, locations and tracks also influence the level of location privacy. They are not included in the metric, because the current version is based on a single snapshot, which is limited in space and time. As a consequence, a track in a snapshot might turn out to be a segment of a trip which ends in a future time, meaning that the complete trip can only be captured with other snapshots.

#### VI. CONCLUSION

In this paper, we have introduced a first approach for quantitatively measuring location privacy of individual users of emerging V2X communication systems. The basic consideration behind is that location privacy of users is not only determined by vehicle tracking, but also by linking vehicle trips to the individuals generated them. Based on snapshots of the V2X system, we capture the information on location privacy in terms of individuals in the system and their trips, which are defined by the origins and destinations of the trips. Assuming that an adversary has information on the linking between vehicles and trips expressed in probabilities, the location privacy of an individual is measured by the uncertainty of such information and quantified as entropy. Then the location privacy of a specific user can be determined by the ratio of

its current entropy and the maximum possible entropy within the given system. The feasibility of the approach is supported by means of different case studies.

In future work we will extend our approach into different directions. First, we plan to incorporate time into our metric by observing and analyzing timely ordered snapshots. Second, we are going to further evaluate our metric on more scenarios and realistic V2X applications. Finally, we will investigate the interrelations between the individual vehicles within a system in order to determine the location privacy of the whole system.

#### REFERENCES

- [1] P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya, Z. Ma, F. Kargl, A. Kung, and J.-P. Hubaux, "Secure vehicular communications: Design and architecture," *IEEE Communications Magazine*, vol. 46, no. 11, pp. 100–109, November 2008.
- [2] F. Dötzer, "Privacy issues in vehicular ad hoc networks," in *Workshop on Privacy Enhancing Technologies*, 2005.
- [3] European Commission, "Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data."
- [4] A. Pfitzmann and M. Köhntopp, "Anonymity, unobservability, and pseudonymity - a proposal for terminology," in *Workshop on Design Issues in Anonymity and Unobservability*, 2000, pp. 1–9.
- [5] S. Steinbrecher and S. Köpsell, "Modelling unlinkability," *Privacy Enhancing Technologies 2003*, pp. 32–47, 2003.
- [6] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *Workshop on Privacy Enhancing Technologies*, 2002.
- [7] C. Diaz, S. Seys, J. Claessens, and B. Preneel, "Towards measuring anonymity," in *Workshop on Privacy Enhancing Technologies*, 2002.
- [8] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 46–55, 2003.
- [9] L. Buttyan, T. Holczer, and I. Vajda, "On the effectiveness of changing pseudonyms to provide location privacy in VANETs," in *ESAS 2007*, July 2007.
- [10] J. Freudiger, M. Raya, M. Felegyhazi, P. Papadimitratos, and J.-P. Hubaux, "Mix-zones for location privacy in vehicular networks," in *WiNITS*, 2007.
- [11] M. Gruteser and B. Hoh, "On the anonymity of periodic location samples," in *Security in Pervasive Computing 2005, Boppard, Germany*, vol. 3450, 2005, pp. 179–192.
- [12] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, and K. Sezaki, "Caravan: Providing location privacy for vanet," in *Proceedings of Embedded Security in Cars (ESCAR)*, Nov. 2005.
- [13] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in GPS traces via density-aware path cloaking," in *ACM Conference on Computer and Communications Security (CCS)*, 2007.
- [14] L. Fischer, S. Katzenbeisser, and C. Eckert, "Measuring unlinkability revisited," in *WPES '08: Proceedings of the 7th ACM workshop on Privacy in the electronic society*, Alexandria, Virginia, October 27 2008.
- [15] A. Dey and G. Abowd, "Towards a better understanding of context and context-awareness," in *Workshop on The What, Who, Where, When, and How of Context-Awareness*, The Hague, The Netherlands, April 2000.
- [16] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House Publishers, 1999.
- [17] J. Krumm, "Inference attacks on location tracks," in *Fifth International Conference on Pervasive Computing*, Toronto, Canada, May 2007, pp. 127–143.
- [18] Z. Ma, F. Kargl, and M. Weber, "Pseudonym-on-demand: a new pseudonym refill strategy for vehicular communications," in *WiVeC 2008*, Calgary, Canada, September 2008.
- [19] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.
- [20] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Enhancing security and privacy in traffic-monitoring systems," *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 38–46, 2006.