Evaluating Location Privacy in Vehicular Communications and Applications

George P. Corser, *Member, IEEE*, Huirong Fu, *Member, IEEE*, and Abdelnasser Banihani

***Abstract*—Vehicular ad hoc networks may one day prevent injuries and reduce transportation costs by enabling new safety and traffic management applications, but these networks raise privacy concerns because they could enable applications to perform unwanted surveillance. Researchers have proposed privacy protocols, measuring privacy performance based on metrics such as k-anonymity. Because of the frequency and precision of location of queries in vehicular applications, privacy measurement may be improved by considering additional factors. This paper defines *continuous network location privacy*; presents KDT-anonymity, which is a composite metric including average anonymity set size, i.e., K, average distance deviation, i.e., D, and anonymity duration, i.e., T; derives formulas to calculate theoretical values of K, D, and T; evaluates five privacy protocols under realistic vehicle mobility patterns using KDT-anonymity; and compares KDT-anonymity with prior metrics.**

***Index Terms*—Location privacy, location based service, LBS, vehicular ad hoc network, VANET, continuous network location privacy, k-anonymity, KDT-anonymity.**

I. Introduction

VEHICULAR ad-hoc networks (VANETs) present uniquely complex location privacy challenges because they utilize protocols which are not used in other networks. For example, in the United States VANET standards are specified by Dedicated Short Range Communications/Wireless Access in Vehicular Environments (DSRC/WAVE), which uses WAVE short message protocol (WSMP). DSRC is designed to send frequent MAC-layer safety messages. DSRC can be used to access location based services (LBS) which also may require frequent precise location (FPL) data.

To what extent can vehicles be protected against unwanted surveillance, even while sending frequent MAC-layer safety messages and accessing LBS applications, both of which requiring FPL data? Let us define this as the *FPL problem*.

Without privacy protections, wireless eavesdroppers, malicious LBS administrators or hackers could track specific vehicles, by cross-referencing vehicles' precise origin and termination coordinates with home and work addresses, using Google Maps or some similar map database, perhaps revealing (*deanonymizing*) a vehicle at a given location at a given time. Because motor vehicles tend to move at high speeds in predictable patterns along prescribed routes, their mobility patterns may make vehicles more vulnerable to location privacy attacks than, say, pedestrian mobile phone users. Deanonymization could occur at either the MAC layer or higher layers. MAC layer VANET systems require vehicles to transmit FPL. LBS applications sometimes have similar requirements. If there is collusion, MAC layer data could be used to circumvent traditional application layer (APP layer) protections, such as spatial-temporal cloaking.

The vehicular location privacy problem is important because driver location data might be misused. Employers might monitor an employee's car arriving at a competitor's office (revealing an employee's job interview) or at specialized medical facilities (revealing an employee's health condition). Eavesdroppers could monitor law enforcement vehicles. It is not difficult to construct potential privacy breaches arising from vehicle surveillance by spouses, ex-spouses, or paparazzi and other stalkers. Legislators have recognized the problem of location privacy. Proposed national level legislation in the United States to address digital location privacy threats includes the Location Privacy Act and the Geolocation Privacy and Surveillance Act.

VANET systems have not yet been widely deployed, but the location privacy problem is not hypothetical as wireless/*wifi* network equipment is being installed in new vehicles. Industry representatives estimate that 90% of vehicles will be *wifi*-connected within the decade [1]. LBS usage continues to grow rapidly [2] and is expected to expand to VANET platforms [3]. Standards governing VANETs [4] provide data fields for future privacy protocols, but the specifics of these protocols remain open research questions. The problem also has special importance in the case of VANETs because the US Department of Transportation (US DOT) has announced it may mandate that all vehicle manufacturers implement VANET equipment on their vehicles. That is, unlike mobile phone users, motorists may not be allowed to remove or turn off their equipment in order to ensure privacy.

To address the vehicle network privacy problem researchers have proposed communication protocols which use the DSRC/WAVE standards. A commonly used metric is k-anonymity. Sweeney introduced the concept of k-anonymity, which requires that in the results of a database query each entity must be indistinguishable from $k-1$ other entities [5]. Gruteser and Grunwald extended k-anonymity to apply to vehicular location privacy [6].

Manuscript received May 3, 2015; revised August 30, 2015 and November 11, 2015; accepted November 21, 2015. The Associate Editor for this paper was Z. Ding.

G. P. Corser is with Oakland University, Rochester, MI 48309 USA, and also with Saginaw Valley State University, University Center, MI 48710 USA (e-mail: gpcorser@svsu.edu).

H. Fu and A. Banihani are with Oakland University, Rochester, MI 48309 USA (e-mail: fu@oakland.edu; abanihani@oakland.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2015.2506579

Contributions: This paper (1) defines *continuous network location privacy*, (2) presents KDT-anonymity, a composite metric including average anonymity set size, K, average distance deviation, D, and anonymity duration, T, (3) derives formulas to calculate theoretical values of K, D and T, (4) evaluates five privacy protocols under realistic vehicle mobility patterns using KDT-anonymity, and (5) compares KDT-anonymity with prior metrics.

The rest of this paper is organized as follows. Section II sets forth the assumptions of the paper, such as system model, threat model and other preliminaries. Section III defines KDT-anonymity and provides equations for theoretical values. Section IV describes simulations, mobility patterns and privacy protocols. Section V evaluates performance of privacy protocols in simulations using KDT-anonymity. Section VI compares KDT-anonymity with existing metrics. Section VII concludes the paper.

II. Preliminaries

This section presents VANET's distinctive communication protocols and system architecture. It also outlines the potential location privacy threats under study in the paper as well as existing privacy metrics and other assumptions.

A. DSRC Protocol Stack

The FCC dedicates a 75 MHz spectrum in the 5.9 GHz band for wireless communication between vehicles. IEEE and SAE have established standards, DSRC/WAVE, to achieve interoperability between devices communicating in this spectrum. The protocol stack features two distinct protocol sets. IPv6/TCP/UDP typically would be used in communication vehicle-to-infrastructure, V2I, such as accessing Internet applications like infotainment or LBSs. WAVE short message protocol, WSMP, would typically be used in communications vehicle-to-vehicle, V2V, such as safety applications. This paper assumes the use of WSMP using security protocols as defined in IEEE 1609.2 and message types defined in SAE J2735. Message sets in SAE J2735 are used to implement privacy protocols. The Basic Safety Message (BSM) is defined in SAE J2735. See Fig. 1.

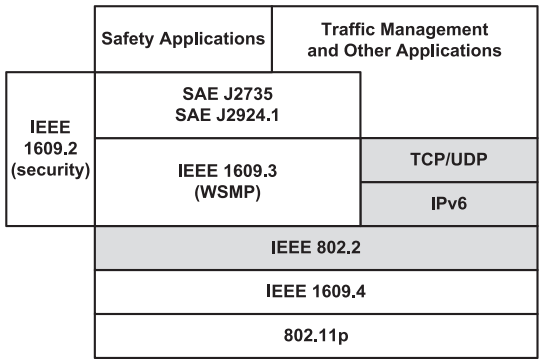

Fig. 1. DSRC protocol stack: includes traditional protocols such as IPv6 and TCP/UDP, but also new protocols such as WAVE Short Message Protocol (WSMP) and SAE J2735.

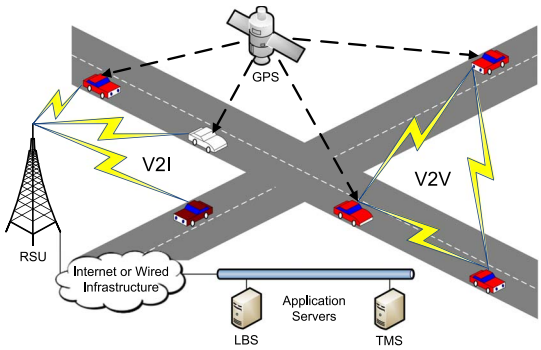

Fig. 2. System model: includes vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. Roadside units (RSUs) relay queries to application servers such as transportation management systems (TMSs).

B. System Model

Vehicular ad hoc networks (VANETs) depend on vehicles each having access to accurate Global Positioning System (GPS) data. Vehicles transmit their positions to each other using vehicle-to-vehicle (V2V) communications. Vehicles may communicate with application servers, such as Location Based Services (LBS) via wired roadside units (RSU) using vehicle-to-infrastructure (V2I) communications. See Fig. 2.

C. Threat Model

Intelligent Transportation Systems (ITSs) are expected to one day include air-traffic-control-like LBS systems called Traffic Management Systems (TMSs) which would enable traffic managers to guide drivers or driverless cars based on current vehicle traffic conditions. If US DOT decides to mandate the use of TMSs, and if vehicles access TMSs through WSMP, then a malicious insider or a outsider who has hacked a TMS might be able to track vehicles from the comfort and anonymity of her own home laptop.

This paper analyzes privacy protocols using the threat modeling framework proposed in [7], which defines the means, actions and goals of the potential attacker. This paper assumes a global passive adversary (GPA) with the following means: *access* to LBS application data, RSU data and perhaps to certain license plate reader (LPR) or other camera data; and *knowledge* of geography (road maps / road topology), traffic conditions (blocked / slow roads), home owner names and addresses and geographical coordinates, and the target's name, address, license plate number, and perhaps expected mobility profile. This paper considers DSRC communications only, not cell phone data or other information from other devices even though that might also be useful to an attacker.

The GPA's actions are assumed to be passive; that is, the GPA would eavesdrop only and would not alter the data being transmitted. The scope of the attack would be global; that is, the GPA could observe data over a wide region, the entire area covered by the TMS. The temporal scope of the actions may be long term; that is, the GPA could eavesdrop for hours, days, months, or even longer periods of time.

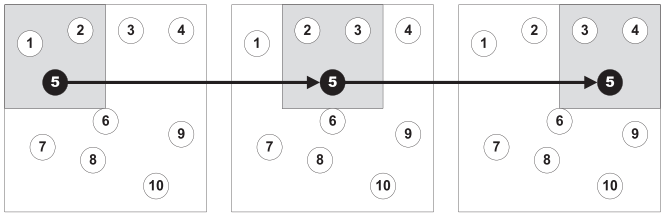

Fig. 3. Spatial cloaking in a series of three snapshots: Vehicle 5 maintains (k, s)-privacy at each snapshot but all three snapshots analyzed together may reveal the vehicle.

The goal of the GPA would be to determine whether a specific vehicle (target) was at a given place at a given time in order to track the target in real time.

D. Why Cloaking Won't Solve the FPL Problem

Spatial cloaking provides effective privacy for LBS users when querying infrequently. If k LBS users are operating in a spatial area, s, then (k, s)-privacy may be achieved [8]. The problem is, if LBS requests are repeated frequently over time, and only one of the k LBS users is consistent throughout the set of cloaked requests, then that user can be exposed. See Fig. 3.

Precision is also an issue. Cloaking depends on obfuscating position. If query results must be precise, then cloaking is ineffective. Further, cloaking requires a trusted third party (TTP) or cloaking proxy. This additional overhead may prove too slow for vehicular LBS applications. Cloaking may also be ineffective in low vehicle densities, for example, if only one vehicle is using the LBS in the given vicinity.

E. Location Privacy Preserving Mechanisms

This paper applies the location privacy framework defined in [7], which enumerates location privacy preserving mechanisms which include four methods: hiding, obfuscating, anonymizing and dummifying.

Previously proposed vehicle network privacy solutions have suggested the following specific methods: silent periods (an example of *hiding*), pseudo identifiers or *pseudoIDs*, mix zones and group signatures (*anonymizing*), spatial and temporal cloaking (*obfuscating*) and transmitting false data, i.e., adding dummy events (this paper calls this *dummifying*).

Consider [7]'s location privacy methods under the vehicular conditions studied in this paper. First, LBS FPL requests and safety communications must use the same precise coordinates because, if LBS and RSU administrators collude, otherwise privacy would be defeated.

Hiding, i.e., going radio silent, would defeat safety and quality of service (QoS). When there is silence neither DSRC safety applications nor LBS will work. If vehicles transmit safety messages but not LBS requests then collusion between LBS and RSU administrators could defeat privacy. TMS is an LBS under control of the same entity that controls RSUs so this would appear possible, even likely.

Obfuscation defeats safety and QoS. Spatial cloaking, for example, would be too slow for safety applications and even if it were fast enough imprecision would degrade safety application effectiveness. Further, if LBS requires FPL for QoS then obfuscating location would cause LBS to return results without the desired precision.

Anonymizing is not enough. PseudoID changing alone, for example, cannot protect against GPAs who track a vehicle over time [10]. Precise location data, even without entity identifiers, perhaps from a series of LBS requests containing FPL, is sufficient to deanonymize a vehicle. If the endpoints are known (e.g., home address latitude/longitude) it becomes especially easy to deanonymize a vehicle [11].

Dummifying defeats safety. False locations could waste resources or even increase the likelihood of crashes. One possible exception is the case of using other vehicles' genuine locations as decoys, as presented in [12]. But even that method requires a silent period for the active user and the decoy to switch pseudoIDs.

In sum, if vehicle privacy is to be preserved then some silence is required, sacrificing some safety and QoS. Some hiding (silent period) must be a part of any privacy protocol hoping to protect against LBS requests with FPL.

F. Traditional Macroscopic Location Privacy Measures

Microscopic location privacy is defined in [7] to be anonymity at a specific place and time. *Macroscopic location* privacy is defined as anonymity from beginning to end of an entire *trajectory*, i.e., a path or segment of a path traversed by a user. This paper examines macroscopic location privacy. Users are vehicles because TMS applications require queries with such rapidity and precision that queries are automated. TMSs therefore are Internet-of-Things (IoT), a.k.a. Internet-of-Cars (IoC) or Internet-of-Vehicles (IoV), applications.

Traditional macroscopic location privacy metrics have included trajectory k-anonymity, entropy of trajectory k-anonymity and tracking probability. Other methods, such as l-diversity [13], t-closeness [14], and (ε, δ)-differential privacy [18], have not enjoyed the same level of popularity and do not always apply to the vehicular FPL problem. This paper adapts the definitions from [15] to define the metrics below.

1) Trajectory k-Anonymity: The anonymity set, AS_i, of target LBS user, i, is the collection of all LBS users, j, including i, within the set of all LBS userIDs, ID, whose trajectories, T_j, are indistinguishable from T_i, within some nonzero probability, p, i.e.,

$$AS_i = \{j | j \in ID, \exists T_j \text{ s.t. } p(i,j) \neq 0\}. \tag{1}$$

2) Entropy of Trajectory k-Anonymity: Entropy represents the level of uncertainty in the correlations between trajectory T_i and trajectories T_j. The entropy H_i of AS_i is as follows. (For more background on this metric, see [25].)

$$H_i = -\sum_{j \in AS_i} p(i,j) \times \log_2 \left(p(i,j)\right). \tag{2}$$

3) Tracking Probability: Tracking probability, Pt_i, is the probability that the anonymity set size of a vehicle's trajectory, $|AS_i|$, is equal to one, which can be written as follows.

$$Pt_i = \Pr\left(|AS_i| = 1\right). \tag{3}$$

If $|AS_i| = 1$, then vehicle i has no anonymity. To measure a system's overall tracking probability, one method is to compute the percentage of vehicles with $|AS_i| = 1$. For example, if 47% of all vehicles have $|AS_i| = 1$, then $Pt = 0.47$ and the system assures 53% anonymity.

4) Limitations of Traditional Measures: One problem with all three above metrics is they do not account for changes in the anonymity set size (*AS-size*) over the course of a vehicle's trajectory. It is possible, even probable, that one or more vehicles in an anonymity set (AS) may become deanonymized, leaving the remaining vehicles in that AS with a reduced *AS-size* during their trajectories. Using only the beginning *AS-size* for all trajectories might *overestimate* the overall anonymity level of a system. Using only the ending anonymity set size might *underestimate* it.

Moreover, the above equations do not quantify spatial dispersion, the span of distance over which vehicles are spread, which *prima facie* seems critical for evaluating *location* privacy. For example, an AS spanning 10 meters may not provide the same location privacy as one spanning 10 miles. Researchers in [19] presented a metric which considers distance deviation over a trajectory, but it measures the distance between locations, not between vehicles, so their equation does not perfectly apply to the conditions considered in this paper.

Finally, the above performance metrics do not consider the duration of anonymity. Privacy protocols perhaps should be considered more effective the longer they last.

III. Proposed Composite Metric

This section presents a composite metric, KDT, for evaluating location privacy provided by privacy-preserving protocols in contexts which involve FPL queries. The composite metric uses anonymity quantification based on k, called k_j, the anonymity set size at a specific point in time, t_j. It uses a Euclidean distance quantification, $\bar{d}_j$, the degree of spatial dispersion, or average span of distance, between members of an anonymity set at a specific point in time, t_j. And it uses a new metric, t_j, anonymity duration, which is the number of contiguous time intervals after t_0 which have anonymity set sizes greater than one. (The first time interval with anonymity set size greater than one is labeled t_0.) The composite metric, kdt, quantifies privacy at a specific point in time. The composite metric, KDT, quantifies the average privacy of a system over a span of time.

A. Definitions of Privacy

Some have defined privacy as user control over information [26], or the degree to which individuals can determine for themselves when, how, and to what extent location information about them is communicated [27]. IEEE 1609.2 measures privacy using anonymity [4]. The term, anonymity, is defined as in set theory. This paper defines privacy, location privacy, network privacy, and continuous privacy as measurable attributes of entities, and measures privacy using anonymity, distance deviation and time.

Definition 1. Privacy: The degree to which an entity cannot be linked to its identity.

Definition 2. Location Privacy: The degree to which a spatial characteristic of an entity cannot be linked to its identity.

Definition 3. Network Privacy: The degree to which an entity cannot be linked to its identity while it is connected to a communications system.

Definition 4. Continuous Privacy: The degree to which, over a contiguous series of time intervals, an entity cannot be linked to its identity.

Definition 5. Continuous Network Location Privacy: The degree to which, over a contiguous series of time intervals, a spatial characteristic of an entity cannot be linked to its identity while it is connected to a communications system.

B. Anonymity Duration $(|T|)$

Anonymity duration, $|T|$, is the size of the set, T, of time intervals, t_j, $j = 0, 1, 2, \ldots, |T|$, over which an entity enjoys continuous network location privacy. So T is a collection of contiguous time intervals starting from t_0 during which the anonymity set size, $|AS_j|$, is greater than one. Formally,

$$T = \{t_j | (t_j = t_0) \vee [(t_{j-1} \in T) \wedge (|AS_j| > 1)]\}. \quad (4)$$

C. Average Anonymity Set Size (K)

Anonymity set size, k_j, is the number of entities that might be confused with one another at time, t_j, that is, $k_j = |AS_j|$. Average anonymity set size, K, is the sum of all k_j from t_0 to t_j divided by $|T| + 1$. Formally,

$$K = \frac{k_0 + k_1 + k_2 + \cdots + k_{|T|}}{|T| + 1}. \quad (5)$$

D. Average Distance Deviation (D)

The distance, d_{sij}, between two entities, s and i, in time interval t_j, is, $d_{sij} = \sqrt{(x_{sj} - x_{ij})^2 + (y_{sj} - y_{ij})^2}$. Let p_{sij} be the probability that an attacker will guess that entity i is the target, given that entity s is the actual target, in time interval t_j. The average distance deviation at time t_j is the weighted sum,

$$\bar{d}_j = \frac{1}{k_j} \frac{1}{k_j} \sum_{s=1}^{k_j} \sum_{i=1}^{k_j} p_{sij} d_{sij}. \quad (6)$$

The average distance deviation over all time intervals is,

$$D = \frac{\bar{d}_0 + \bar{d}_1 + \bar{d}_2 + \cdots + \bar{d}_{|T|}}{|T| + 1}. \quad (7)$$

E. Expected Average Anonymity Set Size $(E[K])$

If a vehicle belongs to an anonymity set as described in (1), and there are $k - 1$ other vehicles in the set, then by definition $|AS| = k$ for all vehicles in AS. To estimate k, assume that vehicles arrive at an anonymization point, P, during a time interval, w, according to a Poisson process, as in [16] and [17]. Let random variable $W = w$ be the fixed time interval during which

vehicles arrive within range of P. Let the inter-arrival time between vehicles have an exponential distribution with a mean of $1/\lambda$. Let X be a random variable, the number of vehicles that arrive within range of P during time W. Then the probability that $X = x$ at $W = w$ can be written as shown in (8). The expected value of X is shown in (9). The expected value of X is the expected value of k.

$$\Pr[X = x|W = w] = \frac{(\lambda w)^x}{x!}e^{-\lambda w} \tag{8}$$

$$E[X|W = w] = \sum_{x=0}^{\infty} x \ \Pr[X = x|W = w] = \lambda w. \tag{9}$$

So, if vehicles arrive at P at a rate of one every five seconds, i.e., $\lambda = 1/5$, in a given 30-second time interval, i.e., $w = 30$, then the expected value of $k = E[k] = E[X] = (1/5)(30) = 6$. By linearity, the expected value of K is the average of all values of k, which is k, i.e., $E[K] = k$.

F. Expected Average Distance Deviation $(E[D])$

It is possible to estimate distance deviation, d, between two vehicles when certain attributes of the roadway topology are assumed, as below.

1) One Point in Time, One Vehicle in Motion $(E[d])$*:* One way to estimate theoretical distance deviation is to identify an anonymization point and compute the expected distance between the anonymization point and some deanonymization point of interest. A unit square region with straight roadways provides mathematically clear illustration and generally correlates with topologies of government jurisdictions and roadways. Assume two vehicles anonymize at the centroid, C, of a square region, R, but that only one of them continues in motion, and that the motion is at a constant rate of speed in a straight line. Assume the moving vehicle has an equal probability of driving from C to any boundary point, B, on the square. What is the expected distance traveled by the moving vehicle? All four sides of the square are line segments of equal size, so C and the endpoints of the line segments form, not only similar triangles, but identical triangles, except for rotation. That is, the average distance from the centroid to one line segment is identical to the average distance from the centroid to any other line segment on the square. Let the square be a unit square centered at (0.5, 0.5) with the x-axis boundary spanning from (0, 0) to (1, 0). The probability density function of a vehicle traveling from the centroid to any point on the x-axis boundary conforms to the uniform distribution.

$$f(x) = \begin{cases} 1, & x \in [0, 1] \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

The expected distance a vehicle would travel after having anonymized at the centroid, C, would be the sum of the products of the probability of traveling from C to a certain boundary point, (x, y), and the Euclidean distance between the centroid and that boundary point. Since y always equals zero on the x-axis boundary the distance from C to (x, y) is

$$dist(C, B) = \sqrt{(x - 0.5)^2 + (0 - 0.5)^2}. \tag{11}$$

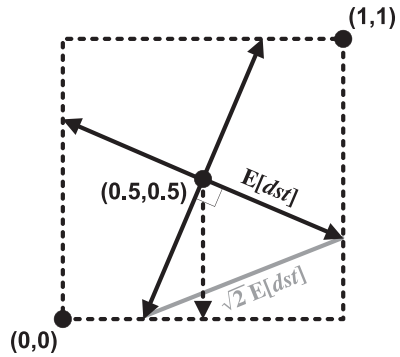

Fig. 4. $E[dst]$ is the expected distance one vehicle would travel from the centroid to the boundary of a square region under certain assumptions. The dotted square represents the region, R, a unit square. The solid perpendicular lines, each starting at (0.5,0.5) and terminating at the boundary of the square, represent the road segments of length $E[dst] = 0.5739$.

To compute the expected distance deviation, sum all possible products by integrating as follows.

$$E[dst] = \int_{-\infty}^{\infty} f(x)dist(C, B)dx$$

$$= \int_{0}^{1} \sqrt{(x - 0.5)^2 + 0.25}\, dx \tag{12}$$

Computing the definite integral above, the expected distance, then, is roughly 0.5739 [20], which is the expected distance deviation from a stationary vehicle remaining at the centroid to a moving vehicle at the time it crosses the boundary of a unit square region.

2) Full Trajectory, Two Vehicles in Motion $(E[d])$*:* What if we wish to know the expected distance deviation, not just at the square's boundary, but over a vehicle's trajectory? With some assumptions it is possible to formulate an estimate. Assume the square region, R, has straight roadways intersecting at perpendicular angles, but that the roadways are not perpendicular or parallel to the sides of the square. Assume the same anonymization point, the centroid, C, which is also the intersection of two straight roads. Assume the distance from the centroid to the endpoint of the straight line road segment is $E[dst]$ as computed in (12). See Fig. 4.

Assume the two members of the anonymity set begin traveling at the same point in time starting at the same intersection. Finally, assume that both vehicles travel at the same rate of speed, r, and there is an equal probability that they will proceed in each of the intersection's four possible directions. Then there are four possibilities. First, if the decoy vehicle travels along the same road as the target vehicle, then $d = 0$ at all points during the trajectory. Second, if the decoy vehicle travels in the opposite direction of the target vehicle, then $d = 2l$, where l is the length of road traveled by one vehicle at any given point in time. Third (and fourth), if the decoy vehicle travels along one of the perpendicular roads, then $d = l\sqrt{2}$, where l is the length of road traveled by one vehicle at any given point in time. There

TABLE I
SIMULATION PARAMETERS

Parameter	Setting
Size of region, R	3000 m x 3000 m
Communications Range	200m, 300 m, or 400m
Mobility Patterns	GMSF (City, Urban, Rural) [21]
Mix Points	50, 60, 70, …, 150, or continuous
Silent Period (Δt)	30 seconds
Simulation Time	2000 s (33.3 min)
Avg. Vehicle Speed	20 m/s

is a 25% probability of each of the four scenarios. The expected value of d, then, would be as follows.

$$E[d] = (0.25)\int_0^{dst}(0 + 2l + l\sqrt{2} + l\sqrt{2})dl$$
$$= (0.25)\int_0^{dst}(2l + 2l\sqrt{2})dl. \quad (13)$$

$E[d]$ evaluates to approximately 0.1988 [20]. From (7) by linearity we can estimate that $E[D] = E[d]$.

G. Expected Anonymity Duration ($E[T]$)

By linearity, the expected anonymity duration is the expected distance of a full trajectory, divided by the constant movement rate, r. Now we can write the following, from (12).

$$E\left[|T|\right] = \frac{E[D]}{r}. \quad (14)$$

IV. SIMULATION

This section describes simulations conducted to evaluate the effectiveness of various privacy protocols using KDT. The experimental set-up is described in Table I.

Traditional privacy solutions obfuscate responses to queries in order to ensure data k-anonymity. This is especially true in the study of trajectory anonymity, as in [28]–[30]. However, in this study all simulations used silent periods to break up the trajectories into disjoint fragments, and this study used active decoy methods, as in [12], to ensure data transmitted by vehicles in an anonymity set were identical except for identifier and location. For example, if there were three vehicles, A, B and C, in an anonymity set, then each vehicle would transmit three queries, for a total of nine queries from the group. We assume the LBS (TMS) permits this in order to prevent giving system administrators or potential hackers the ability to track or trace specific vehicles. The authentication credentials, if necessary, would be encrypted, but the location data would not be encrypted. This has the effect that the LBS administrator would see three vehicles at each of three positions, so she would not know with 100% certainty which vehicle was at which position.

A. Simulation Overview

In each simulation, the goal was to achieve anonymous FPL LBS access, maximizing KDT. That is, vehicles attempted to anonymize with as many other vehicles as possible, remaining anonymous for as great a distance as possible, and for as long a time as possible. Each vehicle did the following.

- Enter region, deanonymized.
- Execute privacy protocol and become anonymized.
- Exit region and become deanonymized again.

The rationale behind the anonymization process outlined above was to create a controlled environment to compare protocols. One could imagine entering a large region where there are few means of ingress and egress, with each ingress/egress point monitored by license plate readers. If a motorist desired to move about the interior of region anonymously, she would have to anonymize after the point of ingress and would lose anonymity upon egress.

B. Mobility Patterns

This research employed Generic Mobility Simulation Framework, GMSF [21], which offers Multi-agent Microscopic Traffic Simulator, MMTS, trace files linked on the GMSF website [22] and provided at the Laboratory for Software Technology website [23], specifically *city*, *urban* and *rural*. All three trace files contain records of time-stamps, vehicle-ids, x-coordinates, y-coordinates within a 3000x3000 meter (9 million square meter) grid. Each mobility pattern starts with a different number of vehicles, v. City starts with $v = 897$. Urban starts with $v = 488$. Rural starts with $v = 110$. Vehicles enter and leave the region at the same rate, but the number of vehicles in the pattern at any given time is not always precisely the same as the number at the start.

Sometimes road topologies, such as in the Freeway pattern (a straight road with perhaps several lanes) and the Manhattan pattern (a grid of horizontal and vertical roads), provide wide ranging linear density versus area density. That is, the vehicle density per linear meter can be out of sync with the vehicle density per square meter, especially when compared with more realistic road topologies. For example, for 900 vehicles in a 3000 × 3000 meter grid, the Freeway pattern might have a linear density of 0.3 v/m, 900 vehicles divided by 3000 meters, and a square density of 0.0001 v/m2, 900 vehicles divided by 9 million square meters. The Manhattan pattern would have a linear density of 0.004839 v/m, 900 vehicles divided by 186,000 meters, but the same square density as the Freeway pattern. In other words, the linear density of the Manhattan pattern is 1.6% that of Freeway pattern given the same square density.

The mobility patterns used in this simulation, however, have similar linear distances: city, 14,783 meters; urban, 13,955 meters; and rural, 10,175 meters. The areas covered are identical, 3000 m × 3000 m, so the mobility patterns provide relatively realistic traffic flows and comparable roadway linear distances and square areas.

C. Location Privacy Protocols and Mix Points

To create mix zones, as defined in [24], the simulation used the concept of a mix point, a position in space and time, with coordinates (x, y, t). The mix point was used to create a circular mix zone of radius, r. If a vehicle was positioned within r, i.e.,

within *comrange*, of (x, y) at time t, then that vehicle initiated whatever the privacy protocol required. If the vehicle never came within comrange of any mix point then it never became anonymized.

This paper evaluates five protocols: SMP-R, stationary mix points, occurring at regular time intervals; SMP-I, stationary mix points, occurring at irregular time intervals; OTFP-R, randomly chosen *on-the-fly* mix points, occurring at regular time intervals; OTFP-I, randomly chosen mix points, occurring at irregular time intervals; and GLRP-2, group leader relay points, which occur continuously throughout the trajectory of a vehicle designated as the leader of a group of vehicles traveling within comrange of the leader. The number, 2, in GLRP-2 indicates that vehicles join the group in pairs. Performance was evaluated against the theoretical values computed in Section III.

1) Stationary Mix Point Protocols (SMP-R and SMP-I): An SMP creates a region that does not move in which vehicles may switch pseudoIDs. A similar protocol is described in [17], but in SMP presented by this paper, a fixed point (x, y) was chosen. To maximize k, the busiest intersection in the mobility pattern was chosen as the "social spot" for mixing. In scenario SMP-R, regular time intervals were chosen. In scenario SMP-I, irregular time intervals were chosen. Vehicles that were within radius, r, of point (x, y) at time, t, were added to the anonymity set. Upon anonymizing, vehicles enter a silent period. They ceased all communications at both MAC and APP layers because, if they were to continue communications via one, under RSU LBS collusion they would be linkable to the other. All vehicles in the anonymity set changed pseudoIDs, but remained silent until the silent period expired, at which point all silent vehicles resumed communications, including anonymous LBS access, using new identifiers.

2) Group Leader Relay Point Protocol (GLRP-2): The group leader protocol has been presented in several important papers, notably [9] and [16]. In this paradigm one vehicle is the designated coordinator, or group leader (GL), of a cluster of vehicles which travel together. In the AMOEBA protocol [16] when two vehicles come within range of a GL they go silent as in SMP. But the mix zone is not stationary. It moves with the GL, centered on the GL. Vehicles must stay within comrange of the GL to communicate anonymously with the LBS.

3) On-the-fly Point Protocols (OTFP-R and OTFP-I): OTFP, similar to the protocol presented in [12], Privacy-by-Decoy (PBD), is similar to SMP except vehicles anonymize at random locations. Timing could be at regular intervals, as OTFP-R, or irregular ones, OTFP-I. If regular time intervals were instituted then there would need to be some method of informing the vehicle as to what the timing would be. If irregular time intervals were instituted, the vehicle could beacon for anonymity at any opportune point along its trajectory. As in the other protocols, in OTFP, when vehicles willing to anonymize move within communications range of each other, they agree to anonymize, go silent for a time, then resume transmissions.

V. Evaluation of Privacy Protocols Using KDT

This section presents the performance of the location privacy protocols using the KDT metric.

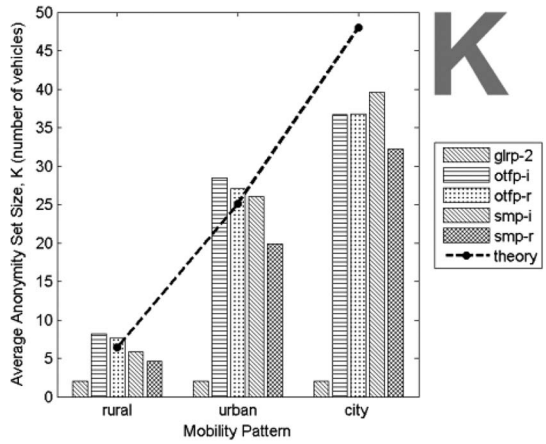

Fig. 5. Average anonymity set size, K, increased with vehicle density. Theoretical K overestimated simulation at higher vehicle densities. Clumpy privacy protocols (OTFP-I, OTFP-R and SMP-I) provided highest K values.

A. Average Anonymity Set Size (K)

The protocols in this study exhibited three noteworthy performance characteristics with respect to anonymity set size. First, as expected, average anonymity set size increased with vehicle density. However, there was a *leveling-off effect*. In the highest density levels, anonymity did not achieve theoretical performance predictions. Average anonymity set size decreased as large proportions of vehicles exited the region early and became deanonymized. When the very last car in an anonymity set exits the region it has a k value of 1. If large numbers of vehicles exit the region early, the remaining vehicles with lower than predicted average k values deflate average anonymity set sizes. See Fig. 5.

Second, the "clumpy" protocols, OTFP-I, OTFP-R and SMP-I, exceeded the performance of the more uniform protocol, SMP-R. (GLRP-2 exhibited small consistent anonymity set size of 2 because vehicles joined in pairs.) Offsetting the leveling-off effect, which decreases k values, clumpiness increases k values. Irregular timing and on-the-fly beacon locations increased clumpiness and thereby improved overall average anonymity. See Fig. 5.

Third, the clumpy protocols were less influenced by the number of mix points. Again, this is because if vehicles are clumped together the number of mix points becomes less relevant. It would be possible for k to equal V, the total number of vehicles, if all vehicles were in the same anonymity set and entered and left the region simultaneously.

B. Average Distance Deviation (D)

Recall that distance deviation is the average length between a vehicle and all of its peers in the same anonymity set. Roadway layout may affect distance deviation performance. Our tests were confined to the GMSF topologies so this report should be understood in that light.

In low densities the on-the-fly protocols outperformed other protocols. In intermediate densities stationary mix points exhibited superior performance. As density increased the protocols performed equivalently. The most important observation from

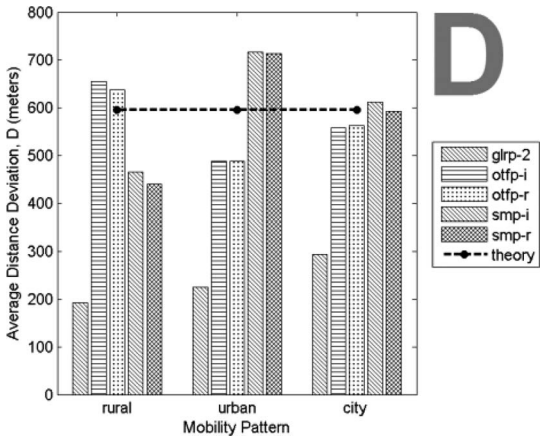

Fig. 6. Distance deviation fluctuated with vehicle density. SMP and OTFP varied in effectiveness at the low, rural, density compared to the medium, urban density. GLRP-2 was lower than others because vehicles must remain within in communications range (300 m) of their group leader at all times.

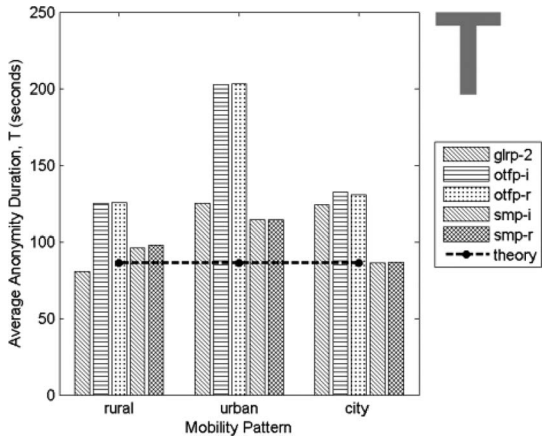

Fig. 7. Anonymity duration remained consistent with increasing vehicle density. On-the-fly protocols outperformed other protocols.

the analysis of distance deviation is that group leader protocol performed noticeably worse than other protocols. This is because vehicles must remain within communications range of the group leader to relay messages. See Fig. 6.

Theoretical/anticipated average distance deviation closely estimated actual results at high densities. There was an unexpected downward trend when evaluating the average distance deviation prediction given the number of mix points in the simulation. This appears to be due to the initial proximity in the anonymity set. With more anonymity sets (more mix points) each initial group started off closer together which decreased the overall average distance deviation as the number of mix points increased.

C. Anonymity Duration (T)

Stationary mix point protocols performed consistently with predictions regarding the anonymity duration of the anonymized vehicles. On-the-fly protocols exceeded predicted performance at all density levels, and the group protocol performed better at higher densities. See Fig. 7.

The reason on-the-fly protocols performed better is because on average they started closer to the center of the region, the optimal point to maximize anonymity duration.

Recall the experiment was set up for vehicles to enter a region, execute a privacy protocol, and then try to maintain anonymity and connectivity before exiting the region. The group leader protocol is similar to the on-the-fly protocols in that it does not use stationary mix points, so GLRP outperformed SMP for the same reason.

In order to maintain relatively constant overall system anonymity, for each density level a stationary mix point was selected based on its ability to anonymize 50% of all vehicles. The region was 3000 $\times$ 3000 m so the centroid was (1500, 1500). However, for SMP privacy protocols the rural mobility pattern set the mix point at (2290,800); the urban mobility pattern at (1430, 2490); and the city mobility pattern at (390,1710). Stationary mix points being closer to the edge, they were more likely to suffer from vehicles exiting earlier and therefore having shorter times of anonymity, reducing overall anonymity duration.

TABLE II
PRIVACY PROTOCOL PERFORMANCE (URBAN MOBILITY PATTERN)

Protocol	K	D	T	k_{max}	H	Pt
GLRP-2	2.00	224.40	125.12	2.00	1.00	0.554
OTFP-I	28.42	488.58	202.69	18.81	4.23	0.602
OTFP-R	27.06	487.63	202.93	19.23	4.27	0.597
SMP-I	26.05	717.57	114.08	20.81	4.38	0.571
SMP-R	19.87	714.39	114.16	24.31	4.60	0.509

D. Summary of Privacy Protocol Performance

Table II shows privacy protocol performance in terms of KDT-anonymity under the urban mobility pattern, specifically, GLRP-2: *low-low-med*, OTFP-I: *high-med-high*, OTFP-R: *high-med-high*, SMP-I: *high-high-low*, SMP-R: *med-high-low*.

The new metric identified clear distinctions between the protocols, except for the OTFP protocols. In fact, the new metric helped reveal the effect of clumpiness on privacy, specifically that privacy protocols need only one method of achieving clumpiness. Using two methods, randomness in both timing and positioning, does not improve performance.

VI. COMPARISON OF KDT WITH PRIOR METRICS

This section compares the performance of the KDT metric with prior metrics, trajectory k-anonymity (k), entropy of trajectory k-anonymity (H) and tracking probability (Pt).

A. K vs. Trajectory k-Anonymity

The value of k as defined in (1) is a vehicle's anonymity at the end of its trajectory. So k for the first vehicle to be deanonymized would be $k_{\max}$, the maximum value of the anonymity set size. This assumes the anonymity set size decreases monotonically, as in the simulation. The *AS-size* of the last vehicle would be 1, assuming vehicles deanonymize one at a time. The average

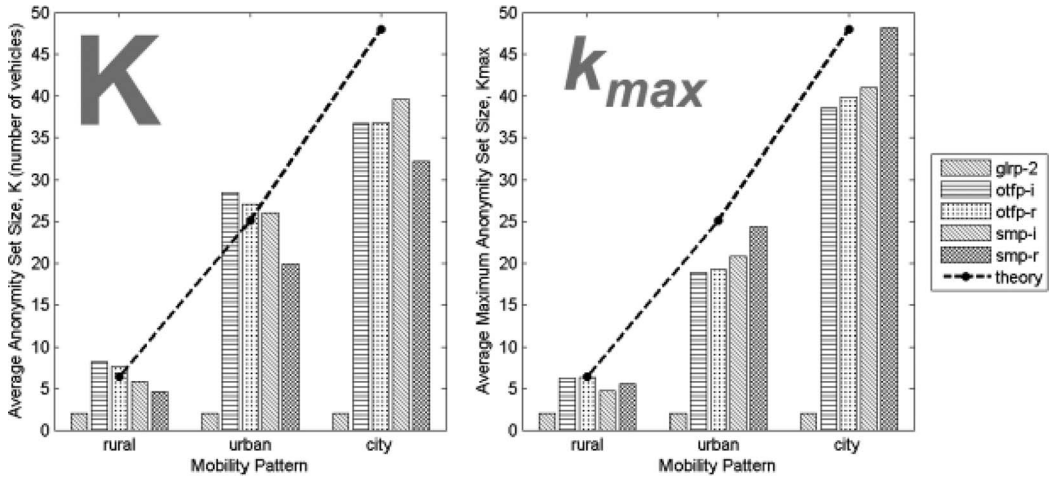

Fig. 8. K vs. trajectory k-anonymity: the latter does not incorporate the effect of higher anonymity in earlier time intervals and consequently reports lower values of k for clumpy protocols and higher values for uniform protocols. The chart at right shows average maximum anonymity set size, $k_{\max}$.

AS-size under these conditions would be $k_{\text{avg}} = (1 + 2 + 3 + \cdots + k_{\max})/k_{\max}$. The new metric, K, averages *AS-size* over many time periods, where the old metric considers only the set of trajectories in the anonymity set at the end of the trajectory.

Why is this important? Because prior to embarking on a trajectory a driver does not know when an attacker will try to deanonymize her. If the attacker investigates historical data and has access to all trajectory information, then the only value of k that matters is the one for the last time interval in the trajectory. However, if a driver wants to choose a privacy protocol that protects her in real time, then it is critical to consider the level of anonymity all along the entire trajectory. This is analogous to the value of car insurance. Looking at the expense historically, if you never had an accident, then car insurance was a waste of money. But looking at the expense as protection against future risks, it may be worthwhile.

Fig. 5 shows the value of K for each privacy protocol. Fig. 8 compares Fig. 5 (left) with the average values of $k_{\max}$ (right), the value of k at the beginning of a trajectory.

B. $H[K]$ vs. Entropy of Trajectory k-Anonymity

In evaluating the relative merits of privacy protocols, the same conclusions apply using K vs. $k_{\max}$, as those using $H[K]$ vs. $H[k_{\max}]$. Researchers in [25] contend that H is preferable to k to quantify the information content in a trajectory. However, this requires knowing ahead of time how many positions are possible in the mobility pattern. Since we cannot know this for certain we use K, rather than $H[K]$.

C. KDT vs. Tracking Probability

Tracking probability is a measure of an entity's chance of having no privacy protection at all. Unlike other metrics, for tracking probability the lower the better. GLRP-2 provided the lowest (best) overall tracking probability, 0.444, a result skewed by performance in rural pattern. All other protocols delivered higher (worse) tracking probabilities than the theoretical prediction, 0.5. See Fig. 9. Clumpy protocols performed worse than SMP-R. In general, the higher (better) the average anonymity set size the higher (worse) the tracking probability, but differences were slight and inconsistent. So the decision to use KDT over Pt may depend on whether a potential target prefers a high level KDT or a low chance of having no protection at all at rural vehicle densities.

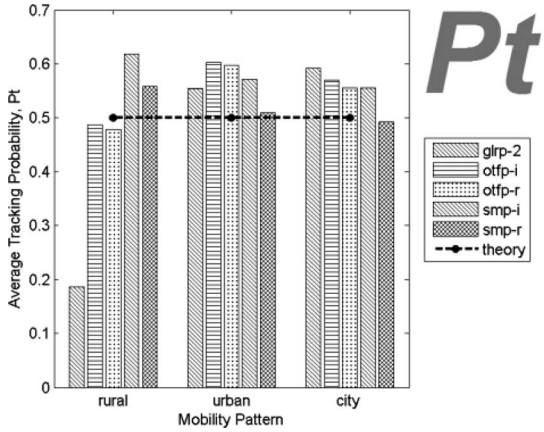

Fig. 9. Tracking probability among privacy protocols was similar because the simulation was deliberately set up to anonymize roughly half of the vehicles. GLRP-2 provided the lowest (best) overall Pt value.

VII. Conclusion

To measure continuous network location privacy what is needed is a metric that answers three questions about VANET privacy protocols. (1) How much anonymity does a protocol provide, given trajectories with fluctuating anonymity set sizes? (2) How much dispersion does a protocol provide, given trajectories with fluctuating distance deviations? (3) How long does a protocol anonymize trajectories? This paper has proposed KDT-anonymity, a composite metric for improving the granularity of the measurement of the property of anonymity over space and time during a trajectory.

Simulation showed how KDT can be used to compare privacy protocols, and how KDT compares with other metrics.

The primary advantage of KDT over prior metrics is that it incorporates changes in the magnitude of anonymity *over time*. While trajectory k-anonymity considers *AS-size* at the end of a trajectory, K is the average *AS-size* over the entire duration of the trajectory. Fig. 8 shows that clumpy privacy protocols have higher K values. This is because more vehicles belong to larger anonymity sets for longer periods of time. This effect is not revealed using trajectory k-anonymity.

Another advantage of KDT is it incorporates the magnitude of dispersion, i.e., distance deviation, over time. Under conditions where a potential target values *how far* she is from her decoys, the D component of KDT becomes a valuable consideration in selecting a privacy protocol.

Finally, if before embarking on a trajectory a potential target values *how long* she may be anonymized, the T component of KDT becomes a valuable consideration in selecting a privacy protocol.

REFERENCES

[1] I. Bush, GM, AT&T Readying in-Vehicle Wi-Fi, Feb. 25, 2013. [Online]. Available: http://philadelphia.cbslocal.com/2013/02/25/gm-att-readying-in-vehicle-wi-fi/

[2] L. Johnson, Location-based services to bring in $4b revenue in 2012: Study, Oct. 31, 2012. [Online]. Available: http://www.mobilemarketer.com/cms/news/research/14115.html

[3] T. Koslowski, Your Connected Vehicle is Arriving, Jan. 3, 2012. [Online]. Available: http://www.technologyreview.com/news/426523/your-connected-vehicle-is-arriving/

[4] *IEEE Standard for Wireless Access in Vehicular Environments Security Services for Applications and Management Messages*, IEEE Std 1609.2-2013 (Revision of IEEE Std 1609.2-2006), Apr. 26, 2013.

[5] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, Oct. 2002.

[6] M. Gruteser and D. Grunwald," Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. 1st Int. Conf. Mobile Syst., Appl. Serv.*, 2003, pp. 31–42.

[7] R. Shokri, J. Freudiger, and J. P. Hubaux, "A unified framework for location privacy," in *Proc. 3rd HotPETs*, 2010, pp. 1–21.

[8] H. Lu, C. S. Jensen, and M. L. Yiu, "Pad: Privacy-area aware, dummy-based location privacy in mobile services," in *Proc. 7th ACM Int. Workshop Data Eng. Wireless Mobile Access*, Jun. 2008, pp. 16–23.

[9] K. Sampigethaya *et al.*, "CARAVAN: Providing location privacy for VANET," in *Proc. Embedded Security Cars*, 2005, pp. 1–15.

[10] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos, "Privacy in inter-vehicular networks: Why simple pseudonym change is not enough," in *Proc. 7th Int. Conf. WONS*, Feb. 2010, pp. 176–183.

[11] G. Corser, H. Fu, T. Shu, P. D'Errico, and W. Ma, "Endpoint Protection Zone (EPZ): Protecting LBS user location privacy against deanonymization and collusion in vehicular networks," in *Proc. 2nd ICCVE*, 2013, pp. 369–374.

[12] G. Corser *et al.*, "Privacy-by-decoy: Protecting location privacy against collusion and deanonymization in vehicular location based services," in *Proc. IEEE Intell. Veh. Symp.*, Dearborn, MI, USA, Jun. 2014, pp. 1030–1036.

[13] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, p. 3, Mar. 2007.

[14] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *Proc. 23rd IEEE ICDE*, Apr. 2007, pp. 106–115.

[15] H. Weerasinghe, H. Fu, and S. Leng, Anonymous service access for vehicular ad hoc networks," in *Proc. 6th Int. Conf. IAS*, Aug. 2010, pp. 173–178.

[16] K. Sampigethaya, M. Li, L. Huang, and R. Poovendran, "AMOEBA: Robust location privacy scheme for VANET," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 8, pp. 1569–1589, Oct. 2007.

[17] R. Lu, X. Li, T. H. Luan, X. Liang, and X. Shen, "Pseudonym changing at social spots: An effective strategy for location privacy in VANETs," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 86–96, Jan. 2012.

[18] F. Kargl, A. Friedman, and R. Boreli, "Differential privacy in intelligent transportation systems," in *Proc. 6th ACM Conf. Security Privacy Wireless Mobile Netw.*, Apr. 2013, pp. 107–112.

[19] T. H. You, W. C. Peng, and W. C. Lee, "Protecting moving trajectories with dummies," in *Proc. Int. Conf. Mobile Data Manage.*, May 2007, pp. 278–282.

[20] [Online]. Available: http://www.numberempire.com/definiteintegralcalculator.php

[21] R. Baumann, F. Legendre, and P. Sommer, "Generic Mobility Simulation Framework (GMSF)," in *Proc. 1st ACM SIGMOBILE Workshop Mobility Models*, May 2008, pp. 49–56.

[22] [Online]. Available: http://gmsf.sourceforge.net/

[23] [Online]. Available: http://www.lst.inf.ethz.ch/research/ad-hoc/car-traces

[24] A. R. Beresford and F. Stajano, "Mix zones: User privacy in location-aware services," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2004, pp. 127–127.

[25] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Comput.*, vol. 2, no. 1, pp. 46–55, Jan.–Mar. 2003.

[26] J. H. Moor, "Towards a theory of privacy in the information age," *ACM SIGCAS Comput. Soc.*, vol. 27, no. 3, pp. 27–32, Sep. 1997.

[27] M. Duckham and L. Kulik, "Location privacy and location-aware computing," in *Dynamic & Mobile GIS: Investigating Change in Space and Time*, vol. 3. Boca Raton, FL, USA: CRC Press, 2006, pp. 35–51.

[28] O. Abul, F. Bonchi, and M. Nanni, "Never walk alone: Uncertainty for anonymity in moving objects databases," in *Proc. 24th IEEE ICDE*, Apr. 2008, pp. 376–385.

[29] M. E. Nergiz, M. Atzori, and Y. Saygin, Towards trajectory anonymization: A generalization-based approach," in *Proc. SIGSPATIAL ACM GIS Int. Workshop Security Privacy GIS LBS*, Nov. 2008, pp. 52–61.

[30] A. Liu *et al.*, Efficient secure similarity computation on encrypted trajectory data," in *Proc. 31st IEEE ICDE*, Apr. 2015, pp. 66–77.

George P. Corser (M'11) received the bachelor's degree in civil engineering from Princeton University, Princeton, NJ, USA, and the master's degree in computer and information sciences from the University of Michigan–Flint, Flint, MI, USA. He is currently working toward the Ph.D. degree with Oakland University, Rochester, MI, USA. He is also an Assistant Professor of computer science and information systems with Saginaw Valley State University, University Center, MI, USA. His current research focuses on vehicular ad hoc network (VANET) security and privacy.

Huirong Fu (M'01) received the Ph.D. degree from Nanyang Technological University, Singapore, in 2000. She is currently a Professor with the Department of Computer Science and Engineering, Oakland University, Rochester, MI, USA, where she joined as an Assistant Professor in 2005. Previously, she was an Assistant Professor with North Dakota State University, Fargo, ND, USA, for three years, and he was a Postdoctoral Research Associate with Rice University, Houston, TX, USA, for two years. As a Lead Professor and the Principal Investigator for several projects funded by the National Science Foundation, she has been actively conducting research in the areas of networks, security, and privacy.

Abdelnasser Banihani received the bachelor's degree in computer information systems from The Hashemite University, Az Zarqa, Jordan, and the master's degree in computer sciences from Jordan University of Science and Technology, Irbid, Jordan. He is currently working toward the Ph.D. degree with Oakland University, Rochester, MI, USA. His current research focuses on vehicular ad hoc network (VANET) security and privacy.