2017 IEEE 37th International Conference on Distributed Computing Systems

Persistent Traffic Measurement Through Vehicle-to-Infrastructure Communications

He Huang†, Yu-E Sun*, Shigang Chen‡, Hongli Xu§, Yian Zhou¶

†School of Computer Science and Technology, Soochow University, Suzhou, China
*School of Urban Rail Transportation, Soochow University, Suzhou, China
‡Department of Computer and Information of Science and Engineering, University of Florida, US
§School of Computer Science and Technology, University of Science and Technology of China, China
¶Google Inc., US
E-mail: sunye12@suda.edu.cn, sgchen@cise.ufl.edu
*Yu-E Sun is the corresponding author.

***Abstract*—Measuring point traffic volume and point-to-point traffic volume in a road system has important applications in transportation engineering. The connected vehicle technologies integrate wireless communications and computers into transportation systems, allowing wireless data exchanges between vehicles and road-side equipment, and enabling large-scale, sophisticated traffic measurement. This paper investigates the problems of persistent point traffic measurement and persistent point-to-point traffic measurement, which were not adequately studied in the prior art, particularly in the context of intelligent vehicular networks. We propose two novel estimators for privacy-preserving persistent traffic measurement: one for point traffic and the other for point-to-point traffic. The estimators are mathematically derived from the join result of traffic records, which are produced by the electronic roadside units with privacy-preserving data structures. We evaluate our estimation methods using simulations based on both real transportation traffic data and synthetic data. The numerical results demonstrate the effectiveness of the proposed methods in producing high measurement accuracy and allowing accuracy-privacy tradeoff through parameter setting.**

***Index Terms*—Vehicular networks, traffic measurement, persistent traffic, privacy.**

I. INTRODUCTION

Measuring traffic volume at points of interest in road systems provides important information for transportation engineering. These point traffic data are useful in estimating traffic link flow distribution as part of investment plan and calculating road exposure rates as part of safety analysis. Much prior research on traffic measurement collects statistics on the number of vehicles passing a certain location during a certain measurement period, often in the form of annual average daily traffic (AADT). Various predication models [1]–[6] have been developed based on the data recorded by roadside units (RSU) installed at road intersections. An example is Mohammed's multiple linear regression model that incorporates demographic variables to measure AADT [1]. Another example is Lam's artificial neural network that estimates AADT based on short period counts [2]. Other research work includes the spatial statistical method by Eom et al. [3], the support vector regression model by Neto et al. [4], the absolute deviation penalty procedure by Yang et al. [5], and the regression and Bayesian model by Tsapakis et al. [6].

The emergence of connected vehicle technologies in the intelligent transportation systems promises radical changes in how transportation traffic measurement will be conducted. The trend is to integrate wireless communications and computers into vehicular cyber-physical systems for better road safety and driving experience [7] [8]. Traffic data collection will become more sophisticated with vehicular communications and networking [9]–[13], such as the Dedicated Short Range Communications standard under IEEE 802.11p [14], which supports wireless data exchanges between vehicles and RSUs.

Such automated systems have been exploited in prior research for collecting point-to-point transportation statistics, i.e., the number of vehicles traveling between any two points (locations) of interest during a certain measurement period in a road system [15], [16]. Point-to-point data provide important input to a variety of transportation studies such as identifying the real sources of traffic congestion and characterizing turning movements at intersections for signal timing determination [17]. There are two performance considerations: The obvious one is the accuracy of traffic measurement. The less obvious one is privacy concern. When vehicles are equipped for wireless communications, there are easy ways to ensure the measurement accuracy. For instance, we may require all vehicles to report their unique IDs to the RSUs that they encounter. In this way, we will be able to figure out the point-to-point traffic volume by comparing the ID sets from two RSUs. However, if a vehicle keeps transmitting its ID to RSUs, its entire moving history is recorded in great details. Such large-scale, universal tracking of movement raises privacy concern [15], [16], [18].

In this paper, we take a step further to study a new problem of persistent traffic volume measurement. After we measure the point traffic volume at a certain location over time for many measurement periods, we naturally want to mine the data for more knowledge. Given a certain number of measurement periods, the point persistent traffic is defined as the set of

DOI 10.1109/ICDCS.2017.327

vehicles that pass the location in all those periods. The rest is treated as transient traffic. For example, we may want to learn the persistent traffic volume over the workdays of a week, over the Saturdays of several weeks, or on all days in a month. Such data tells us the amount of core, stable traffic at a location, as the transient traffic varies over time. Similarly, after measuring the point-to-point traffic volume between two locations for many measurement periods, we want to know the persistent traffic volume containing common vehicles that show up during each period from one location to the other. For example, if a location is consistently congested, we can find the sources of the traffic based on point-to-point traffic measurement. Yet the persistent point-to-point traffic measurement tells us the minimum amount of traffic contribution that we can always expect from each of those sources. This information helps in determining the priority order for planning measures of traffic relief at various sources.

The problem of point (point-to-point) persistent traffic measurement is challenging if we want to achieve both measurement accuracy and privacy protection (which prevents even the authority from learning the trajectories of the moving vehicles). We propose two novel estimators for privacy-preserving persistent traffic measurement: one for point traffic and the other for point-to-point traffic. The basic idea is for each RSU to encode the vehicles passing by during each measurement period in a privacy-preserving data structure, called traffic record, where the information of all vehicles is mixed such that the vehicle identities are hidden. To estimate the point persistent traffic, we join the traffic records produced during the periods of interest and derive an estimation formula based on the join result through probabilistic analysis. Similarly, to estimate the point-to-point persistent traffic, we first join the traffic records at each location and then join their results, from which we derive an estimation formula. We evaluate our estimation methods using simulations based on both real transportation traffic data and synthetic data. The extensive simulations demonstrate the effectiveness of the proposed methods in producing persistent traffic estimation of high accuracy and allowing accuracy-privacy tradeoff through parameter setting.

II. Preliminaries

A. Persistent Traffic Measurement

We study an intelligent transportation system with vehicle-to-infrastructure communication capability. Road-Side Units (RSUs) are deployed at locations of interest, such as street intersections. All RSUs are connected wirelessly or by wire to a central sever, where data are collected and processed for transportation traffic management functions. Each vehicle also has a unique ID and is equipped to communicate with the RSUs through DSRC [14].

Traffic measurement is performed in each measurement period (e.g., a day), whose length is set as needed by the authority. During an arbitrary period, each RSU records the passing vehicles in a privacy-preserving data structure, called *traffic record*, without keeping any identifying information such as vehicle IDs. We study the following problems.

First, consider a single location L and a set Π of traffic records produced from the RSU at L during a number of measurement periods — for example, records from Monday through Friday of a certain week, records from Mondays of three consecutive weeks, or several records of interest based on any other criterion. A *common vehicle* refers to a vehicle that passes location L in all the measurement periods of interest. All common vehicles form the *persistent traffic*. The first problem, called *point persistent traffic measurement*, is to use the traffic records in Π to estimate the volume of persistent traffic, i.e., the number of common vehicles passing L.

Next, consider two locations, L and L'. Let Π be a set of traffic produced by the RSU at L during a number of measurement periods of interest, and Π' be the set of traffic records produced by the RSU at L' during the same measurement periods. With respect to two locations, a *common vehicle* refers to a vehicle that passes both locations in all the measurement periods of interest, and accordingly the *point-to-point persistent traffic* is the aggregate of such vehicles. The second problem, called *point-to-point persistent traffic measurement*, is to use Π and Π' to estimate the volume of point-to-point persistent traffic, i.e., the number of common vehicles that pass both L and L'.

B. Security and Threat Model

Vehicles will only interact with RSUs from trustworthy authorities. This can be easily enforced through authentication based on PKI. Communications begin with an RSU broadcast beacons, each carrying its public-key certificate, which was obtained from a trusted thirty party and was pre-installed with the RSU. When a vehicle receives a beacon, it uses its pre-installed public key of the trusted third party to verify the certificate. If not successful, the vehicle will keep silent; otherwise, it performs authentication with the RSU using the latter's public key obtained from the verified certificate. After successful authentication, it performs vehicle recording with the RSU as will be explained in the next subsection, with all data exchanges encrypted. Rogue RSUs may be deployed by non-authorities; they will fail the authentication with the vehicles, which will reject further communications.

We assume a semi-trusted model for the authorities. The transportation authority has good faith in implementing the proposed privacy-preserving methods since their goal is not to track people, but only to gather transportation traffic, which provides input for city development planning (without any real-time or short-term consequences). Their RSUs will communicate with the passing vehicles and perform all operations as expected. However, as the traffic records are produced and stored. At a later time, other people (such as police or FBI) who gain access to the records may exploit the information to track individual vehicles when they have the need to do so. For instance, if a hypothetical system design requires every vehicle to transmit its unique identifier to each encountered RSU, then these recorded identifiers may be used to track the trajectory of any vehicle. In order to prevent this from happening, it is highly desired that a vehicle will not transmit its unique ID, nor transmit any other fixed number to the RSUs that it passes.

Moreover, we assume that an anonymous MAC protocol such as SpoofMAC [19] is used to support privacy preservation such that the MAC address of a vehicle is not fixed. With such a protocol, before a vehicle communicates with an RSU, it picks a temporary MAC address randomly from a large space for one-time use, which prevents the MAC address from serving as an identifier of the vehicle.

C. Performance Metrics

We consider the following two performance metrics to evaluate persistent traffic measurement.

1. *Estimation Accuracy*: Let n_* be the actual volume of persistent traffic, i.e., number of common vehicles passing one location (or two locations) during the measurement periods of interest. Let $\hat{n_*}$ be the volume estimated based on the traffic records. We measure the estimation accuracy by evaluating the relative error, $\frac{|\hat{n_*}-n_*|}{n_*}$. A good traffic measurement method is expected to have close-to-zero relative errors.

2. *Preserved Privacy*: The essence of privacy preservation in transportation traffic measurement is to allow the tracker only a limited chance to identify any part of the trajectory of any vehicle. Following [15], [16], we want to make sure that anyone that possesses the traffic records cannot definitively determine any trace of any vehicle. In general terms, the traffic records may indicate that a vehicle has passed from one location to another location when the vehicle actually did not, and the records may indicate that the vehicle has not passed from one location to another location when the vehicle actually did.

As we will see shortly, the traffic records are probabilistically constructed. Consider a vehicle v and suppose we somehow know that the vehicle has passed a location L. The vehicle never transmits its identifier, but there could be some external ways that reveal its presence at a certain location — for example, the vehicle is stopped by a police car for speeding at the location. Now, the question of concern by this paper is how much additional information the traffic records will leak to reveal the trajectory of v. Consider another arbitrary location L', let p be the probability that the traffic records will show that v has passed both locations even though v did not; clearly, p represents a noise term that the design of traffic record introduces. Let p' be the probability that the traffic records will show that v has passed both locations when v actually did so; p' includes the noise contribution p. We will later derive the formulas for p and p'. The privacy protection is weak if the probabilistic noise p approaches to zero while the probabilistic information $p' - p$ approaches to one, which implies near-definitive tracking. We want to increase the former, relative to the latter. If we can increase the noise p to a level that is comparable to or even outweights $p' - p$ by far, there will be increasingly significant doubt in what the traffic records indicate. Therefore, we use $\frac{p}{p'-p}$, called the probabilistic noise-to-information ratio, to characterize the level of privacy protection. We expect this ratio to be at least greater than one, and the larger the better.

D. Traffic Record and Vehicle Encoding

Consider an arbitrary RSU installed at a certain location and an arbitrary measurement period. The data structure of traffic record is a bitmap B of m bits. Each vehicle that passes the RSU during the period is encoded by a bit, which is pseudo-randomly selected from B in a way that masks the identity of the vehicle yet leaves a probabilistic signature, allowing statistical analysis for traffic volume. The size of B, i.e., the value of m, may differ at different RSUs or at different measurement periods for the same RSU. We will come back to set m later.

The basic observation is that there is a functional relationship between the number of ones (or zeros) in B and the number of vehicles encoded — the more the number of vehicles is, the more the number of ones in B will be. Based on that function, we can estimate the number of vehicles from the number of ones. The problem of persistent traffic measurement will be more difficult as we need to combine the information from the traffic records of different periods to figure out the number of common vehicles, which we discuss in the next two sections. Below we define how the traffic record B is constructed in each measurement period. In order to support privacy, we want to mix the information from different vehicles in B. The vehicle-encoding method should have the following properties: (1) When vehicles are encoded at a certain location, different vehicles may be probabilistically encoded by the same bit. (2) When a vehicle passes multiple locations (RSUs), it may be encoded at different bit indices. Together, they break the one-to-one association between vehicles and bits.

The traffic record is constructed as follows: At the beginning of each measurement period, the bits in B are reset to zeros. The RSU broadcasts beacons in preset intervals, such as once per second, ensuring that each passing vehicle will be able to receive a beacon, which carries the RSU's location L, its public-key certificate, and the size m of its bitmap. After a vehicle receives a beacon, it verifies the certificate and uses the public key to authenticate the RSU. After verifying that the RSU is from a trusted authority, the vehicle computes the following hash output: $h_v = H(v \oplus K_v \oplus C[H(L \oplus v) \mod s]) \mod m$, where H is a hash function that provides good randomness, v is the vehicle ID, K_v is a private key known only by the vehicle, L is the location of the RSU, and C is an array of s randomly selected constants. Because h_v is a function of L, its value may be different at different locations; the system parameter s controls the number of different values that h_v may take; the use of randomized constants in C helps improve the quality of input to the outer hash. The vehicle transmits h_v to the RSU, which will in turn set the bit at index h_v to one, i.e., $B[h_v] = 1$. That is the only operation of vehicle encoding. At the end of each measurement period, the RSU will send the content of the bitmap B as its traffic record to the central server, where queries may be submitted from the users to estimate point or point-to-point persistent traffic.

The index h_v produced from a vehicle is not predictable by others because the private key K_v is not known. Moreover, the array C of constants are also known only to the vehicle. During a measurement period, many vehicles may pass an RSU. Due to vehicles' random selection of bits to set, different vehicles may choose the same bit as a result of hash collision. The same vehicle may choose different indices at different locations because the hash output is also dependent on the location L. Such mixing and variation in vehicle encoding help preserve privacy and make it harder for a tracker (including the authority) to definitively determine the trajectory of any vehicle.

Let $h_v(i) = H(v \oplus K_v \oplus C[i]) \mod m$, where $1 \leq i \leq s$. We call $B[h_v(i)]$, $1 \leq i \leq s$, the *representative bits* of vehicle v in bitmap B. When the vehicle passes the RSU, it selects one of the representative bits uniformly at random through another hashing, $i = H(L \oplus v) \mod s$. The size s of the array C determines the number of different representative bits from which a vehicle may choose to set. As our privacy analysis and numerical evaluation will show, this parameter controls a performance tradeoff between preserved privacy and traffic estimation accuracy.

From each bitmap B reported by an RSU, the central server can estimate the number of vehicles passing the RSU during the corresponding measurement period based on linear probabilistic counting [20]–[22] as follows:

$$\hat{n} = -m \ln V_0, \tag{1}$$

where V_0 is the fraction of bits in B that are zeros. Based on the historic traffic volumes, the central server will set the bitmap size at each RSU as follows:

$$m = 2^{\lceil \log_2(\bar{n} \times f) \rceil}, \tag{2}$$

where $\bar{n}$ is the expected traffic volume at the RSU during the measurement period based on historical average at the same location and the same time, and f is a system-wide load factor that specifies the ratio of the bitmap size and the expected traffic volume.

Formula (1) allows us to estimate point traffic based a single traffic record B. But we cannot apply it directly to solve the new problems of measuring *persistent* point traffic and *persistent point-to-point* traffic across multiple traffic records. The key issues are how to combine multiple traffic records and how to derive new estimation formulas based on the combined information. Because the traffic volume $\bar{n}$ varies from place to place, the bitmap size varies accordingly. We set the value of m in (2) always as a power of two in order to facilitate joining the information of different bitmaps for persistent traffic estimation; such joining will become clear when we discuss the technical details.

III. Measurement of Point Persistent Traffic

Given a set of t bitmaps, $\{B_1, ..., B_t\}$, that are measured at a certain location L of interest during t measurement periods,

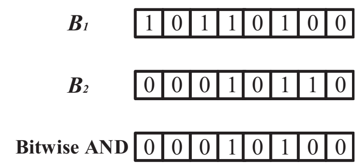

Fig. 1: An example of combining two bitmaps of the same size, B_1 and B_2, by bitwise AND

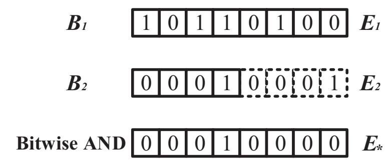

Fig. 2: An example of combining two bitmaps of different sizes by bitwise AND

we want to estimate the point persistent traffic over the set as defined in Section II-A. Let m be the largest size of all bitmaps, i.e., $m = \max\{l_1, ..., l_t\}$, where l_j is the number of bits in B_j, for $1 \leq j \leq t$.

A. Joining the Bitmaps through Expansion

To find the common traffic encoded by multiple bitmaps, we need to join the information from the bitmaps. Recall that each vehicle is encoded by setting a bit. If all bitmaps have the same size, one simple approach of combining them is to perform bitwise AND, as shown by an example in Fig. 1. If a bit in the resulting bitmap is one, it means the same bit must be one in all bitmaps B_1 through B_t, indicating that there may be a common vehicle setting the bit in all those measurement periods.

However, if the bitmaps have different sizes, we will not be able to perform bitwise AND directly among them. To circumvent this problem, if the size of a bitmap B_j is smaller than m, we expand it by replicating it multiple times until its size reaches m, as shown by an example in Figure. 2, where B_2 is replicated once (dashed part). Such expansion is always possible because the sizes of all bitmaps are powers of 2. The expanded bitmap is denoted as E_j. If $l_j = m$, then E_j is simply B_j. We use Π to denote the set of expanded bitmaps (also known as traffic records). We perform bitwise AND over all expanded bitmaps in Π, and the result is denoted as E_*. Its ith bit is denoted as $E_*[i]$, $1 \leq i \leq m$.

Consider an arbitrary common vehicle v. Its hash output, $h_v = H(v \oplus K_v \oplus C[H(L \oplus v) \mod s]) \mod m$, gives the index of the bit in E_* that the vehicle is mapped to. In order to make sure that all common vehicles are recorded by E_*, the following property should hold after bitwise AND: $E_*[h_v] = 1$, for any common vehicle v. It is obvious that this property holds in the special case where all original bitmaps B_j, $1 \leq j \leq t$, have the same size m. In the general case, thanks to the design fact that the bitmap sizes are two's powers, we prove the property as follows: Consider an arbitrary bitmap B_j of

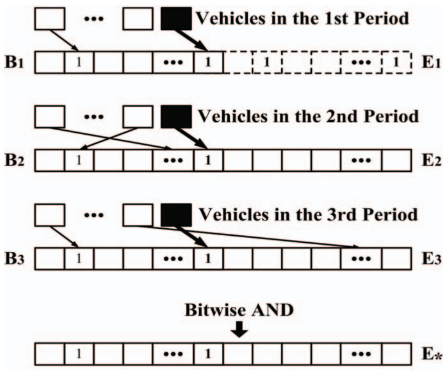

Fig. 3: An illustration for vehicle encoding in bitmaps, bitmap expansion, and bitmap joining

size l_j. The bit set to one by v is at index h_v mod l_j. After l_j is expanded to m, all the bits in E_j at indices $(h_v \mod l_j)+kl_j$, $0 \le k < \frac{m}{l_j}$, are ones. Because both l_j and m are powers of 2 and $m \ge l_j$, we know that $\frac{m}{l_j}$ is a positive integer. Hence, $h_v \mod m = (h_v \mod l_j) + k'l_j$, for a certain integer $k' \in [0, \frac{m}{l_j})$. Therefore, the bit in E_j at index (h_v mod m) must be one. Since this holds for all expanded bitmaps $E_j, 1 \le j \le t$, we conclude that $E_*[h_v] = 1$.

Can we simply estimate the number of common vehicles based on the number of ones in E_*? The answer is no because transient vehicles can also cause bits in E_* to be ones. See Figure 3 for example, which shows three bitmaps, B_1, B_2 and B_3, collected from the same location. The vehicles that appear in each measurement period are shown above each bitmap. A black box indicates a common vehicle that appears in all measurement periods at the location. A white box indicates a transient vehicle. Each vehicle sets a bit to one, as the arrows in the figure show. The size of B_1 is half of the other bitmaps' size. We expand B_1 to E_1 by doubling its size, as shown in the figure with dashed lines. E_*, which is the bitwise AND of the three bitmaps, is at the bottom of the figure. We only show the values of two bits in E_*; both are ones. The first bit of one is caused by transient vehicles, which are different cars but happen to set bits at the same index due to hash collision. The second bit of one is caused by a common vehicle. We also want to point out that two common vehicles may set the same bit to one due to hash collision. Therefore, a bit of one in E_* may indicate zero, one or multiple common vehicles. Estimating common vehicles solely based on ones in E_* will be inaccurate. But if we combine information in Π into more than one bitmap and use that information jointly with E_*, we will be able to gain enough differentiation through probabilistic derivation, which in turn allows us to make meaningful estimation.

B. Deriving an Estimator for Persistent Traffic

We divide Π into two subsets, $\Pi_a = \{E_1, ..., E_{\lceil t/2 \rceil}\}$ and $\Pi_b = \{E_{\lceil t/2 \rceil+1}, ..., E_t\}$. Let E_a be the join of bitmaps in Π_a by bitwise AND, E_b the join of bitmaps in Π_b by bitwise AND, and E_* the join of E_a and E_b by bitwise AND.

E_a (or E_b) encodes both the set of common vehicles and possibly some transient vehicles. From (1), we compute the number of independent vehicles that would have produce the bitmap E_a (or E_b):

$$n_a = \frac{\ln V_{a,0}}{\ln(1-\frac{1}{m})}, \qquad n_b = \frac{\ln V_{b,0}}{\ln(1-\frac{1}{m})} \tag{3}$$

where $V_{a,0}$ ($V_{b,0}$) is the fraction of zeros in E_a (E_b). Essentially we use an abstract set of n_a vehicles to produce the same effect as what all vehicles in Π_a jointly produce in E_a. This abstraction relieves us from the dependency within Π_a. Because the bits of ones in E_a retain the information from the common vehicles, the n_a vehicles contain the set of common vehicles. Similarly we use an abstract set of n_b vehicles to summarize the effect of Π_b. While dividing Π into more than two sets is possible, we find the two-set solution is not only simple but works effectively.

For an arbitrary bit in E_*, its value can be modeled as a random binary variable whose value is probabilistically determined as the vehicles randomly choose their bits to set. Let n_* be the number of common vehicles. The probability P_* for at least one of the common vehicles to set the bit is

$$P_* = 1 - (1 - \frac{1}{m})^{n_*}. \tag{4}$$

The probability for this bit to be set by a transient vehicle in E_a (or E_b) is

$$P_a = 1 - (1 - \frac{1}{m})^{n_a - n_*}, P_b = 1 - (1 - \frac{1}{m})^{n_b - n_*}. \tag{5}$$

Let $X_{i,1}$, $1 \le i \le m$, be the event that the ith bit in E_* becomes one. Combining the above analysis, the probability for $X_{i,1}$, $1 \le i \le m$, to occur is

$$\begin{aligned} &Prob\{X_{i,1}\} = P_* + (1 - P_*)P_aP_b \\ &= 1 - (1 - \frac{1}{m})^{n_*} + (1 - \frac{1}{m})^{n_*} \times (1 - (1 - \frac{1}{m})^{n_a - n_*}) \times \\ &\qquad (1 - (1 - \frac{1}{m})^{n_b - n_*}) \\ &= 1 - (1 - \frac{1}{m})^{n_a} - (1 - \frac{1}{m})^{n_b} + (1 - \frac{1}{m})^{n_a + n_b - n_*}. \end{aligned} \tag{6}$$

Transforming (3) to $V_{a,0} = (1 - \frac{1}{m})^{n_a}$ and $V_{b,0} = (1 - \frac{1}{m})^{n_b}$, we have

$$Prob\{X_{i,1}\} = 1 - V_{a,0} - V_{b,0} + V_{a,0}V_{b,0}(1 - \frac{1}{m})^{-n_*} \tag{7}$$

Let $V_{*,1}$ be a random variable for the fraction of bits in E_* that are ones. We can measure an instance value of $V_{*,1}$ from E_*. This instance value will be used in the estimator derived later. We have

$$V_{*,1} = \frac{1}{m} \sum_{i=1}^{m} I_{X_{i,1}}, \tag{8}$$

where $I_{X_{i,1}}$ be the indicator variable of $X_{i,1}$, whose value is 1 when the event $X_{i,1}$ occurs and 0 otherwise. Clearly, $E(I_{X_{i,1}}) = Prob\{X_{i,1}\}$. Hence,

$$E(V_{*,1}) = \frac{1}{m} \sum_{i=1}^{m} E(I_{X_{i,1}}) = \frac{1}{m} \sum_{i=1}^{m} Prob\{X_{i,1}\}. \tag{9}$$

Because $Prob\{X_{i,1}\}$, $1 \leq i \leq m$, has the same value in (7), we have

$$E(V_{*,1}) = 1 - V_{a,0} - V_{b,0} + V_{a,0}V_{b,0}(1 - \frac{1}{m})^{-n_*}. \quad (10)$$

Solving the equation for n_*, we have

$$n_* = \frac{\ln V_{a,0} + \ln V_{b,0} - \ln(E(V_{*,1}) + V_{a,0} + V_{b,0} - 1)}{\ln(1 - \frac{1}{m})}. \quad (11)$$

Replacing the expected value $E(V_{*,1})$ with the instance value $V_{*,1}$ measured from E_*, we have the following formula for an estimated value $\hat{n_*}$ of the number of common vehicles.

$$\hat{n_*} = \frac{\ln V_{a,0} + \ln V_{b,0} - \ln(V_{*,1} + V_{a,0} + V_{b,0} - 1)}{\ln(1 - \frac{1}{m})}. \quad (12)$$

where $V_{a,0}$, $V_{b,0}$ and $V_{*,1}$ are measured from E_a and E_b and E_*, respectively.

IV. Measurement of Point-to-Point Persistent Traffic

Consider two locations of interest, L and L'. Let $\{B_1, ..., B_t\}$ and $\{B'_1, ..., B'_t\}$ be the sets of bitmaps measured during the same periods at L and L', respectively. We want to estimate the point-to-point persistent traffic between the locations as defined in Section II-A. Let m (m') be the largest size of all bitmaps from L (L'). Without loss of generality, assume $m \leq m'$.

A. Two-Level Bitmap Expansion and Joining

The first level of bitmap expansion and joining are performed among the bitmaps from a single location. Consider the bitmaps from L. For each bitmap B_j, $1 \leq j \leq t$, if its size is smaller than m, we expand it by replicating it multiple times until its size reaches m. We then perform bitwise AND over all expanded bitmaps from L. The resulting bitmap is denoted as E_*, whose size is m. As we have explained in Section III-A, the bitmap E_* encodes the set C of common vehicles appearing at one location L during t measurement periods. Besides that, E_* also encodes transient vehicles, e.g., those vehicles that set the first bit of one in E_* in Figure 3.

Similarly, we expand each bitmap from L' to the size of m' and perform bitwise AND over all expanded bitmaps from L'. The result is denoted as E'_*, which encodes the set C' of common vehicles appearing at L' during the t measurement periods, as well as transient vehicles. What we are interested here is not $|C|$ or $|C'|$; they are the subject of the previous section. Let $C'' = C \bigcap C'$. We want to know $|C''|$, the number of common vehicles that pass both L and L' during the t measurement periods. When we discuss the point-to-point common vehicles in C'', the vehicles in C or C'' but not in C'' will also be referred to as transient vehicles.

The second level of bitmap expansion and joining are performed between two locations. If $m < m'$, we expand E_* by replicating it multiple times until its size research m'. The expanded bitmap is denoted as S_*, where we use S to signify this is the Second level expansion. If $m = m'$, S_* is simply E_*. We join the expanded S_* with E'_* by bitwise OR and the resulting bitmap is denoted as E''_*. (The reason for bitwise OR instead of bitwise AND is that the probabilistic analysis for deriving an estimator based on the result of bitwise AND is extremely difficult, whereas bitwise OR gives a closed-form formula.)

B. Deriving an Estimator for Point-to-Point Persistent Traffic

In Section III-B, a common vehicle always sets bits in E_a and E_b at the same index, which makes probabilistic analysis much simpler. For persistent point-to-point traffic measurement, a common vehicle may set bits in E_* and E'_* at difference indices, $h_v = H(v \oplus K_v \oplus C[H(L \oplus v) \mod s]) \mod m$ and $h'_v = H(v \oplus K_v \oplus C[H(L' \oplus v) \mod s]) \mod m'$, which are dependent on location coordinates, L and L'. This makes the problem much harder because a common vehicle does not necessarily set bits in E_* and E'_* at the same index. It only has a certain probability to do so.

E_* (or E'_*) encodes both the set C'' of common vehicles and possibly some transient vehicles. Again based on (1), we compute the number of independent vehicles that would have produced the bitmap E_* (or E'_*):

$$n = \frac{\ln V_{*,0}}{\ln(1 - \frac{1}{m})}, \qquad n' = \frac{\ln V'_{*,0}}{\ln(1 - \frac{1}{m'})} \quad (13)$$

where $V_{*,0}$ ($V'_{*,0}$) is the fraction of zeros in E_* (E'_*). Similar to Section III-B, we use an abstract set of n independent vehicles to produce the same effect as what all vehicles that pass L will jointly produce in E_*. Yet the bits of ones in E_* retain all information from the common vehicles. We also use an abstract set of n' independent vehicles to summarize what the vehicles passing $L's$ will produce in E'_*.

For an arbitrary bit $E''_*[i]$, $1 \leq i \leq m'$, whose value is the OR of $S_*[i]$ and $E'_*[i]$. We derive the probability for $E''_*[i]$ to be zero. For this to happen, no common/transient vehicle should set $S_*[i]$ or $E'_*[i]$ to one. Let n'' be the number of common vehicles in C''.

First, consider an arbitrary common vehicle in C''. If the vehicle sets $E_*[i \mod m]$ to one, then after expansion $S_*[i]$ will be one. Let $E_*[i' \mod m]$, $1 \leq i' \leq m'$, be the bit that the vehicle sets at L. The probability for $E_*[i' \mod m]$ to be different from $E_*[i \mod m]$ is $1 - \frac{1}{m}$. In this case, $S_*[i]$ is not set by the vehicle. Under this condition ($i' \mod m \neq i \mod m$), we analyze the probability of the vehicle not setting $E'_*[i]$ at location L'. The vehicle will be mapped at L' to one of its s representative bits, including $E'_*[i']$ with probability $\frac{1}{s}$ — in which case, $E'_*[i]$ is not set because $i' \neq i$. With probability $1 - \frac{1}{s}$, the vehicle is mapped to a bit other than $E'_*[i']$, and that bit has a chance of $\frac{1}{m'}$ to happen to be $E'_*[i]$. In summary, the probability for any common vehicle not to set either $S_*[i]$ or $E'_*[i]$ is $(1 - \frac{1}{m})(\frac{1}{s} + (1 - \frac{1}{s})(1 - \frac{1}{m'}))$. The probability for none of the common vehicles to set either $S_*[i]$ or $E'_*[i]$ is

$$P_1 = (1 - \frac{1}{m})^{n''}(\frac{1}{s} + (1 - \frac{1}{s})(1 - \frac{1}{m'}))^{n''}. \quad (14)$$

Second, there are $n - n''$ transient vehicles passing location L. The probability for none of these transient vehicles to set

$E_*(i \mod m)$ is $(1-\frac{1}{m})^{n-n''}$. Similarly, there are $n'-n''$ transient vehicles passing location L'. The probability for none of these transient vehicles to set $E'_*[i]$ is $(1-\frac{1}{m'})^{n'-n''}$.

Now we model the value of the ith bit in E''_* as a binary random variable, where $1 \le i \le m'$. Let $Y_{i,0}$, $1 \le i \le m'$, be the event that the ith bit remains zero. Combining the above analysis, we have

$$\begin{aligned} Prob\{Y_{i,0}\} &= P_1(1-\frac{1}{m})^{n-n''}(1-\frac{1}{m'})^{n'-n''} \\ &= (1+\frac{1}{sm'-s})^{n''}(1-\frac{1}{m})^{n}(1-\frac{1}{m'})^{n'} \end{aligned} \tag{15}$$

Applying (13), we have

$$Prob\{Y_{i,0}\} = (1+\frac{1}{sm'-s})^{n''}V_{*,0}V'_{*,0} \tag{16}$$

Let $V''_{*,0}$ be a random variable for the fraction of bits in E''_* that are zeros. We have

$$V''_{*,0} = \frac{1}{m'}\sum_{i=1}^{m'} I_{Y_{i,0}}, \tag{17}$$

where $I_{Y_{i,0}}$ be the indicator variable of $Y_{i,0}$, whose value is 1 when the event $Y_{i,0}$ occurs and 0 otherwise. Clearly, $E(I_{Y_{i,0}}) = Prob\{Y_{i,0}\}$. Hence,

$$E(V''_{*,0}) = \frac{1}{m'}\sum_{i=1}^{m'} E(I_{Y_{i,0}}) = \frac{1}{m'}\sum_{i=1}^{m'} Prob\{Y_{i,0}\}. \tag{18}$$

Because $Prob\{Y_{i,0}\}$, $1 \le i \le m'$, has the same value in (16), we have

$$E(V''_{*,0}) = (1+\frac{1}{sm'-s})^{n''}V_{*,0}V'_{*,0}. \tag{19}$$

Suppose m' is large. Solving the equation for n'' and apply $\ln(1+x) \approx x$ when x is small, we have

$$n'' \approx sm'(\ln E(V''_{*,0}) - \ln V_{*,0} - \ln V'_{*,0}). \tag{20}$$

Replacing the expected value $E(V''_{*,0})$ with the instance value $V''_{*,0}$ measured from E''_*, we have the following formula for an estimated value $\hat{n''}$ of the number of common vehicles passing both locations L and L' during all t measurement periods.

$$\hat{n''} = sm'(\ln V''_{*,0} - \ln V_{*,0} - \ln V'_{*,0}). \tag{21}$$

where $V_{*,0}$, $V'_{*,0}$ and $V''_{*,0}$ are measured from E_* and E'_* and E''_*, respectively.

V. Privacy Analysis

When a vehicle passes an RSU, the only thing that a vehicle does is to set a bit in the RSU's bitmap to one at an index that may vary from location to location. Moreover, different vehicles may choose the same indices. What each RSU gathers is a bitmap, with each bit of one suggesting the passage of at least one vehicle. Therefore, the tracker may possibly identify the trajectory of a common vehicle through the observation that bits with the same index at two different locations are both ones. Below, we analyze privacy preservation of our persistent-traffic measurement design in terms of the probabilistic noise-to-information ratio as defined in Section II-C.

When a vehicle v passes a location L, it sends an index value i to the RSU, which set the bit at the index in the bitmap B to one, i.e., $B[i]=1$. Let m be the size of B. Suppose the authority is able to associate the index i with the vehicle v at L, for example, when the vehicle is stopped by a police for speeding, there is no other vehicle around, and the police informs the authority. Now if the authority finds at a different location L' that the bit at the same index in the bitmap B' is also one, i.e., $B'[i]=1$, can it assert that the vehicle v has moved from L' to L, thus revealing the partial trajectory of the vehicle?

Recall that other vehicles may choose the same index and the same vehicle may choose difference indices at different locations. The bit $B'[i]$ may have been set by other vehicles passing L'; in this case, the above assertion about the trajectory of v will be wrong. Let p be the probability that $B'[i]$ is set to one by other vehicles even if v does not pass L'. Let n' be the number of vehicles passing L', each having a probability of $\frac{1}{m'}$ to set $B'[i]$. Therefore,

$$p = 1-(1-\frac{1}{m'})^{n'}. \tag{22}$$

Let p' be the probability that $B'[i]$ is set to one when v does pass L'. According to Section II-D, the same vehicle may set bits at different indices at different locations. In particular, it has s representative bits and randomly selects one to set at L'. Therefore, the probability for v to set $B'[i]$ to one is $\frac{1}{s}$. We know by (22) that other vehicles will set $B'[i]$ with probability p. Hence,

$$p' = p + (1-p)\frac{1}{s}. \tag{23}$$

The probabilistic noise-to-information ratio is therefore

$$\frac{p}{p'-p} = \frac{1-(1-\frac{1}{m'})^{n'}}{(1-\frac{1}{m'})^{n'}\frac{1}{s}}. \tag{24}$$

VI. Simulation

In this section, we perform simulations to evaluate the performance of our proposed persistent traffic estimators in terms of estimation accuracy and preserved privacy under different parameter settings. We use both real transportation traffic and synthetic traffic. To the best of our knowledge, this is the first work that studies persistent traffic measurement (as defined in Section II-A) through vehicle-to-infrastructure communications. There are prior privacy-preserving approaches for measuring point-to-point traffic [15], [16] or measuring travel time [23]. But there is no prior work that measures *persistent* point-to-point traffic under the same model in Section II-B. We stress these are very different problems. Therefore, we will compare the proposed estimators with some benchmark methods of simpler designs to demonstrate the effectiveness of the proposed design.

A. Simulation Results Based on Real Traffic Data

First, we use the real-world vehicle trip table measured at the city of Sioux Falls, South Dakota; the data can be found in [24], which contains the actual traffic volume from one point to another in the city. In our simulation, we generate the point-to-point common vehicles between two locations L and L' based on the number $n^{''}$ from the vehicle trip table, and then randomly generate $n - n^{''}$ transient vehicles for L and $n' - n^{''}$ transient vehicles for L', where n (n') is the total traffic volume, i.e., the sum of all entries in the trip table involving L (L'). The bitmap size m (m') is computed from n (n') and f; see Section II-D. In the simulation, we let L' be the location with the largest total traffic volume of all, with $n' = 451000$. We randomly select 8 other locations as L.

We simulate 10 measurement periods with randomly generated transient vehicles. The performance of our estimator on point-to-point persistent traffic between L and L' is shown in Table I, where we set $s = 3$ and $f = 2$. The results are the average of 1000 simulation runs. It can be seen from the fourth row that $m' \neq m$ and their ratio ranges from 2 to 16. The 6th-9th rows present the relative error (defined in Section II-C) when $t = 3, 5, 7, 10$, respectively. In Table I, we can see that the estimation error is mostly small. The error is higher when $L = 8$, where the number of common vehicles is just 3,000, comparing with 451,000 vehicles passing L' and 28,000 vehicles passing L; in this case, the noise generated from the transient vehicles is high, relative to the number of common vehicles.

We include the last line in the table for a benchmark comparison with a simpler design where we set $m' = m$ and m is determined by n and f, which is to ensure the privacy of the vehicles pass location L. Everything else stays the same as described in the paper. The relative error in the last line is larger than that in the 7th line (which is also bolded); in both lines, $t = 5$. For example, when $n'' = 3,000$ in the last column, the relative error of the proposed estimator is just 0.0585, whereas the relative error of the same-size design is 1.3749.

The Sioux Falls data can only support limited evaluation. We resort to synthetic data for other simulations.

B. Simulation Results Based on Synthetic Traffic

Next, we evaluate the proposed estimators based on synthetic traffic data. For point persistent traffic measurement, the number of vehicles that passes L during each measurement period is randomly generated from the range of $(2000, 10000]$. Let n_{min} be the minimum number of generated vehicles that pass location L in any measurement period. We set the number of common vehicles n_* at L during all measurement periods from $0.01n_{min}$ to $0.5n_{min}$, with steps of $0.01n_{min}$. We set $s = 3$ and $f = 2$. We compare the proposed estimator (Section III) with a benchmark method of a simpler design that estimates directly from E_* with $\hat{n}_* = \frac{lnV_{*,0}}{ln(1-1/m)}$ [20]–[22], where E_* is the bitwise AND of all t bitmaps from L.

The simulation results are presented in Fig. 4, where the horizontal axis represents the actual persistent traffic volume and the vertical axis represents the relative error. The left plot is the comparison between the proposed estimator and the benchmark when $t = 5$, the right plot is the comparison when $t = 10$. In both cases, the proposed estimator significantly outperforms the benchmark, particularly when the persistent traffic volume is relatively small. The relative error becomes much smaller when t is increased from 5 to 10. That is because the AND join of more bitmaps helps filter out the ones produced by transient vehicles (which are noise).

For point-to-point persistent traffic measurement, the number of vehicles that passes L (or L') is randomly generated from $(2000, 10000]$, and thus the two locations have the same average traffic. Suppose n_{min} (n'_{min}) is the minimum number of generated vehicles that passed L (L') during any measurement period. Let $n''_{min} = \min\{n_{min}, n'_{min}\}$. We set the number $n^{''}$ of common vehicles from $0.01n''_{min}$ to $0.5n''_{min}$, with step size of $0.01n''_{min}$.

Under different values of f, Fig. 5 - 6 present the measurement accuracy in a different form, with each point in the figures representing a measurement, where the x-coordinate of the point is the actual persistent traffic volume and the y-coordinate is the estimated volume. We also draw the equality line $y = x$. The closer the points are to the line, the better the measurement accuracy will be. As the points cluster around the equality line, two plots is each figure confirm that the proposed estimators produce good measurement accuracy for both point persistent traffic and point-to-point persistent traffic. When we increase the value of f from 2 to 3, the estimation accuracy is visibly better. Recall that f is the ratio of the bitmap size and the expected traffic volume. Increasing f means that the bitmap size is increased, which reduces the mixing of information from different vehicles, thus improving accuracy, but in the meantime reducing privacy protection, as we will see next.

C. Preserved Privacy

In Table II, we examine privacy protection by measuring the probabilistic noise-to-information ratio with respect to f and s. We know that the larger this ratio is, the better the privacy protection will be, because it will become increasingly uncertain to use the traffic records to track individual vehicles. We want this ratio to be at least greater than 1. We see from the table that the ratio increases when f decreases or s increases. Earlier we have observed that the estimation accuracy moves in the opposite direction: it decreases when f decreases or s increases. So there is a tradeoff between accuracy and privacy. Based on all our numerical evaluations, we believe $f = 2$ and $s = 3$ make a good compromise between the two. Under these parameters, our accuracy evaluation has consistently produced good results, and the probabilistic noise-to-information ratio is about 2 as shown in the table. In the last row, we also give the noise probability p that the traffic records will show a vehicle passes both locations even when it actually does not. The value of p only depends on f. It is about 40% when $f = 2$. A noise-to-information ratio of 2 implies a probability of 60% that the traffic records will show a vehicle passes both locations when

L	1	2	3	4	5	6	7	8
n	213000	140000	121000	78000	76000	47000	40000	28000
m	524288	524288	262144	262144	262144	131072	131072	65536
m'/m	2	2	4	4	4	8	8	16
n''	40000	20000	19000	8000	8000	7000	6000	3000
relative error $(t = 3)$	0.0122	0.0167	0.0210	0.0369	0.0361	0.0398	0.0438	0.0948
relative error $(t = 5)$	**0.0101**	**0.0144**	**0.0169**	**0.0252**	**0.0267**	**0.0284**	**0.0265**	**0.0585**
relative error $(t = 7)$	0.0111	0.0151	0.0171	0.0257	0.0241	0.0279	0.0251	0.0518
relative error $(t = 10)$	0.0104	0.0139	0.0172	0.0258	0.0256	0.0261	0.0234	0.0497
same-size bitmaps $(t = 5)$	**0.0110**	**0.0172**	**0.0267**	**0.0510**	**0.0491**	**0.1271**	**0.1305**	**1.3749**

TABLE I: relative error of point-to-point persistent traffic volume estimation in the Sioux Falls network

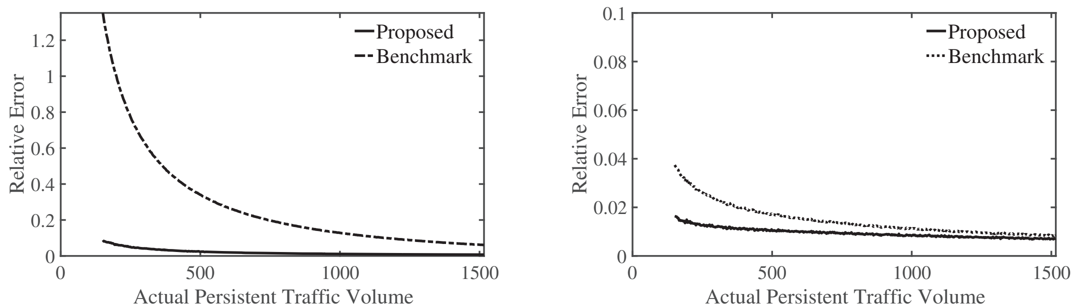

Fig. 4: Relative error of point persistent traffic estimation. *Left plot*: $t = 5$; *right plot*: $t = 10$.

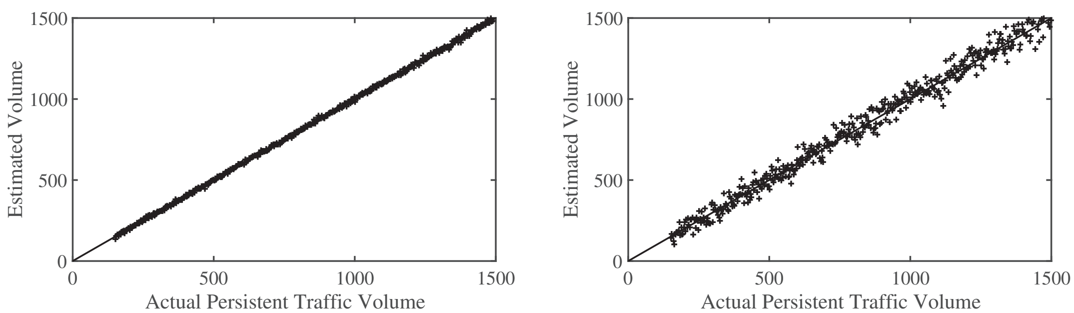

Fig. 5: *Left plot:* measurement accuracy of point persistent traffic volume ($t = 5, f = 2$); *right plot:* measurement accuracy of point-to-point persistent traffic volume ($t = 5, f = 2$).

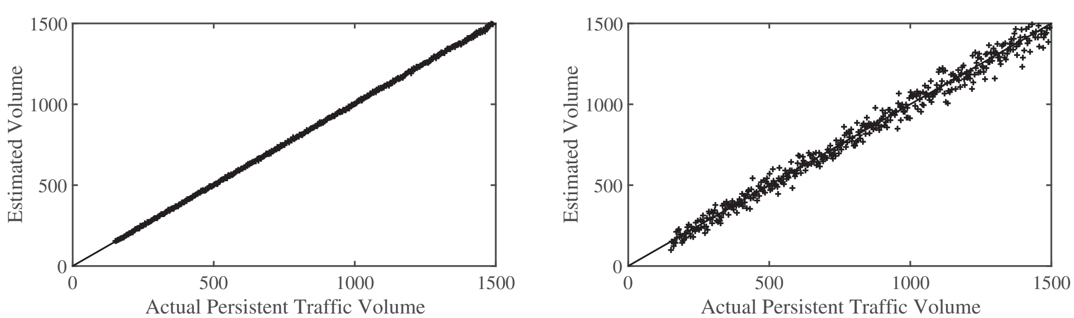

Fig. 6: *Left plot:* measurement accuracy of point persistent traffic volume ($t = 5, f = 3$); *right plot:* measurement accuracy of point-to-point persistent traffic volume ($t = 5, f = 3$).

s \ f	$f=1$	$f=1.5$	$f=2$	$f=2.5$	$f=3$	$f=3.5$	$f=4$
$s=2$	3.4368	1.8956	1.2975	0.9837	0.7912	0.6614	0.5681
$s=3$	5.1553	2.8433	1.9462	1.4755	1.1869	0.9922	0.852
$s=4$	6.8737	3.7911	2.5950	1.9673	1.5825	1.3229	1.1361
$s=5$	8.5921	4.7389	3.2437	2.4592	1.9781	1.6536	1.4201
p	0.6321	0.4866	0.3935	0.3297	0.2835	0.2485	0.2212

TABLE II: Privacy preserving: the probabilistic noise-to-information ratio and noise p

it does, including the noise contribution of 40%. Noise (40%) overwhelms information (20%) by a ratio of 2 to 1, making any tracking result very questionable.

VII. Conclusion

This paper studies the new problems of persistent point traffic measurement and persistent point-to-point traffic measurement in the context of intelligent vehicular networks, where the vehicles can communicate with the RSUs wirelessly. We present the operation protocol that the RSUs use to encode the vehicles in their traffic records. We propose two novel estimators for measuring point persistent traffic volume and point-to-point persistent traffic volume. The estimator design considers both measurement accuracy and privacy preservation. We analyze the preserved privacy of the estimators. The numerical evaluation demonstrates the effectiveness of the proposed methods in producing high measurement accuracy and allowing accuracy-privacy tradeoff through parameter setting.

Acknowledgement

The research of authors is partially supported by National Science Foundation (NSF) CNS-1409797, STC-1562485, National Natural Science Foundation of China (NSFC) under Grant No. 61572342, No. 61672369, Natural Science Foundation of Jiangsu Province under Grant No. BK20151240, No. BK20161258, China Postdoctoral Science Foundation under Grant No. 2015M580470, No. 2016M591920. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of author(s) and do not necessarily reflect the views of the funding agencies (NSF, and NSFC).

References

[1] D. Mohamad, K. C. Sinha, T. Kuczek, and C. F. Scholer, "Annual Average Daily Traffic Prediction Model for County Roads," *J. of the Transportation Research Board*, 1998.

[2] W. Lam and J. Xu, "Estimation of AADT from Short Period Counts in Hong Kong – A Comparison Between Neural Network Method and Regression Analysis," *J. of Advanced Transportation*, pp. 249–268, 2000.

[3] J. K. Eom, M. S. Park, T. Heo, and L. F. Huntsinger, "Improving the Prediction of Annual Average Daily Traffic for Nonfreeway Facilities by Applying a Spatial Statistical Method," *J. of the Transportation Research Board*, pp. 20–29, 2006.

[4] M. C. Neto, Y. Jeong, M. K. Jeong, and L. D. Han, "AADT Prediction using Support Vector Regression with Data-Dependent Parameters," *Expert Systems with Applications*, vol. 36, pp. 2979–2986, March 2009.

[5] B. Yang, Y. Wang, S. Wang, and Y. Bao, "Efficient Local AADT Estimation via SCAD Variable Selection Based on Regression Models," *Control and Decision*, pp. 1898–1902, 2011.

[6] I. Tsapakis, W. H. Schneider, and A. Nichols, "A Bayesian Analysis of the Effect of Estimating Annual Average Daily Traffic for Heavy-Duty Trucks using Training and Validation Data-Sets," *Transportation planning and technology*, pp. 201–217, 2013.

[7] J. Eriksson and H. Balakrishnan, "Cabernet: Vehicular Content Delivery Using WiFi," *Proc. of MOBICOM*, 2008.

[8] U. Lee, J. Lee, J. Park, and M. Gerla, "FleaNet: A Virtual Market Place on Vehicular Networks," *IEEE Trans. on Vehicular Technology*, 2010.

[9] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected Vehicles: Solutions and Challenges," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 289–299, 2014.

[10] M. Pan, P. Li, and Y. Fang, "Cooperative Communication Aware Link Scheduling for Cognitive Vehicular Ad-hoc Networks," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 30, no. 4, pp. 760–768, 2012.

[11] J. Sun, C. Zhang, Y. Zhang, and Y. Fang, "An Identity-based Security System for User Privacy in Vehicular Ad Hoc Networks," *IEEE Trans. on Parallel and Distributed Systems (TPDS)*, vol. 21, no. 9, pp. 1227–1239, 2010.

[12] Y. Zhu, Y. Wu, and B. Li, "Vehicular Ad Hoc Networks and Trajectory-Based Routing," *Internet of Things*, pp. 143–167, 2014.

[13] X. Zhu, S. Jiang, L. Wang, and H. Li, "Efficient Privacy-Preserving Authentication for Vehicular Ad Hoc Networks," *IEEE Trans. on Vehicular Technology*, vol. 63, no. 2, pp. 907–919, 2014.

[14] Y. L. Morgan, "Notes on DSRC & WAVE Standards Suite," *IEEE Comm. Surveys & Tutorials*, 2010.

[15] Y. Zhou, Z. Mo, Q. Xiao, S. Chen, and Y. Ying, "Privacy-Preserving Transportation Traffic Measurement in Intelligent Cyber-Physical Road Systems," *IEEE Transactions on Vehicular Technologies*, vol. 65, pp. 3749–3759, May 2016.

[16] Y. Zhou, S. Chen, Z. Mo, and Q. Xiao, "Point-to-Point Traffic Volume Measurement through Variable-Length Bit Array Masking in Vehicular Cyber-Physical Systems," *Proc. of IEEE ICDCS*, 2015.

[17] USDOT, "Traffic Monitoring Guide," 2013. [Online]. Available: http://www.fhwa.dot.gov/policyinformation/tmguide

[18] L. Kong, L. He, X.-Y. Liu, Y. Gu, M.-Y. Wu, and X. Liu, "Privacy-preserving compressive sensing for crowdsensing based trajectory recovery," in *Proc. of INFOCOM*, 2015, pp. 31–40.

[19] SpoofMAC, "Spoof your MAC address," 2015. [Online]. Available: https://github.com/feross/SpoofMAC

[20] K.-Y. Whang, B. T. Vander-Zanden, and H. M. Taylor, "A Linear-time Probabilistic Counting Algorithm for Database Applications," *ACM Transactions on Database Systems*, vol. 15, no. 2, pp. 208–229, June 1990.

[21] C. Estan, G. Varghese, and M. Fish, "Bitmap Algorithms for Counting Active Flows on High-Speed Links," *IEEE/ACM Trans. on Networking*, vol. 14, no. 5, October 2006.

[22] M. Yoon, T. Li, S. Chen, and J. Peir, "Fit a Compact Spread Estimator in Small High-Speed Memory," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1253–1264, October 2011.

[23] P. Fuxjaeger, S. Ruehrup, T. Paulin, and B. Rainer, "Towards privacy-preserving wi-fi monitoring for road traffic analysis," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 3, pp. 63–74, 2016.

[24] L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla, "An efficient approach to solving the road network equilibrium traffic assignment problem," *Transportation Research*, vol. 9, no. 5, pp. 309–318, 1975.