# Short Paper: A Practical View of "Mixing" Identities in Vehicular Networks

Bisheng Liu
School of Computer Science
Fudan University
220 Handan Road, Shanghai, China 200433
bsliu@fudan.edu

Jerry T. Chiang, Jason J. Haas, Yih-Chun Hu
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
1406 W. Green Street, Urbana, IL 61801-2918
{chiang2, jjhaas2, yihchun} @illinois.edu

## ABSTRACT
In a Vehicular Ad hoc NETwork (VANET), vehicles broadcast safety messages disclosing their trajectory information in order to warn drivers of impending accidents. Precise location information needed for these safety applications, combined with the need to exclude attackers through the use of authentication, creates a significant privacy risk. One method proposed to improve privacy is the use of many pseudonyms, and changing pseudonyms while in a *mix zone* where all other vehicles also change pseudonyms. Previous work has evaluated the effectiveness of mix zones using traces generated based on traffic theory. In this paper, we analyze the privacy obtainable from using mix zones in VANETs based on actual recordings of vehicle movements. We choose *rank* instead of *entropy* as our privacy metric because, as we will show, entropy is difficult to measure in our scenarios.

## Categories and Subject Descriptors
C.2.0 [**Computer-Communication Networks**]: General—Security and protection, (e.g., firewalls)

## General Terms
Security, Performance

## Keywords
Vehicular ad hoc networks, privacy, mix zones

## 1. INTRODUCTION
Researchers have proposed using Vehicular Ad hoc NETworks (VANETs) to disseminate safety messages to support vehicular safety applications, which they hope will improve drivers' safety. For example, the US Department of Transportation (US-DOT) has identified eight such applications [1]. Most of these safety applications require that each vehicle periodically broadcast its current location among other information (e.g., speed, timestamp, etc.) to its one-hop neighbors. Depending on the safety application, each vehicle may be required to broadcast safety messages at a rate of up to 10 Hz. While these applications could improve vehicular safety, an attacker can also try to track a single vehicle by eavesdropping on the unsuspecting vehicle's safety messages.

In order to preserve *anonymity* and ensure *untraceability*, each vehicle could use randomly-changing and unlinkable pseudonyms to communicate with each other. Pseudonyms are sufficient for initiating communication since most safety applications are more concerned about vehicle trajectory than vehicle identity. However, using the temporal and spatial relationship between the current and previous locations of each vehicle, an attacker capable of monitoring all communications in the network could make pseudonym changes ineffective. To reduce an attacker's ability to correlate multiple pseudonyms from a single vehicle, the vehicle can choose to change its pseudonyms only in regions where the attacker is unable to monitor the communications, known as *mix zones* [2]. Several researchers [3][4] have suggested that mix zones in VANETs be created at predetermined locations where the density of vehicles is high and the speed and direction of vehicles change often, such as at intersections.

The obtainable privacy from a given mix zone depends not only on the sampled traffic traces crossing the mix zone, but also on the power of the tracking algorithm chosen by the attacker. Prior work [3][5] has often relied on simulating VANET environments to evaluate the effectiveness of mix zones, using traffic data generated by simulators based on traffic theory. To the best of our knowledge, we are the first to apply a realistic traffic model as a part of the analysis of the achieved privacy from a mix zone based on *real vehicle mobility data*. For a mix zone created at a busy intersection, we extend previously proposed tracking algorithms by taking traffic signals and lane changes into consideration; for mix zones created on a high-density straight road, we choose a heuristic tracking algorithm that correlates vehicles based on the sequence of vehicles entering and exiting the road section. Because an attacker that lacks a more sophisticated traffic model can always use our proposed tracking algorithm, our evaluations represent an *upper bound* on the amount of privacy that the mix zone under evaluation can provide. Moreover, prior research has used *entropy* [3] to measure the privacy obtainable from a mix zone based on a particular tracking algorithm. However, prior work calculates entropy using probabilities determined by a tracking algorithm, and may not be a very useful metric in evaluating privacy, as we will discuss in Section 3. Our research addresses the following questions. *How much privacy can we obtain from a mix zone in reality? How do we measure the obtained privacy?*

The remainder of this paper is organized as follows. In Section 2 we present related work, our assumptions, and our attacker model. In Section 3, we discuss our choice of privacy metrics. In Section 4, we present the tracking algorithm we use for mix zones created at road intersections and analyze the obtainable privacy. We then present the tracking algorithm we use for mix zones along straight roads and analyze the effectiveness of these mix zones in Section 5. Finally we conclude our work in Section 6.


*WiSec'11*, June 14–17, 2011, Hamburg, Germany.

# 2. BACKGROUND

## 2.1 Related Work

The privacy implications of VANETs are a major concern [6][7]. Researchers have proposed that vehicles should change pseudonyms while communicating in order to mitigate the threat of being tracked by an attacker [6][8][9]. However, an attacker might still be able to determine that two pseudonyms correspond to a single vehicle by using the information included in safety messages, such as the location and the speed of the vehicle, thereby compromising the driver's privacy.

Many authors have suggested using *silent periods* [10] to preserve anonymity; that is, when vehicles need to update their pseudonyms, they turn off their transceivers for a period of time. Sampigethaya et al. [11][12] proposed that vehicles in geographic proximity can form a group and elect a group leader so that the group leader can communicate on behalf of the whole group. Other group members could then safely extend their silent periods. The same researchers also developed a user-centric approach, *Swing & Swap* [13], to increase location privacy by loosely synchronizing pseudonym updates between vehicles. As we mentioned in Section 1, since many safety applications require each vehicle to periodically broadcast safety messages approximately every 100 milliseconds, the maximum silent period is bounded by the required broadcast interval.

Beresford et al. [2][14] proposed using *mix zones* for privacy preservation. Mix zones are locations where an attacker is unable to directly link vehicular communications to vehicle identities, and hence, the vehicle identities could be mixed in such areas. If many vehicles simultaneously update pseudonyms in a mix zone, an attacker is less likely to be able to associate each updated pseudonym to its previous pseudonym. Freudiger et al. [3] proposed that mix zones could be created at road intersections and used simulations to determine the effectiveness of their mix zones. Their approach is only effective against external attackers that have no legitimate credentials, because an attacker with legitimate credentials can understand all transmissions within the mix zone, thus able to track a vehicle from mix zone entry to exit. Other approaches have considered connecting several mix zones together to form a mix network that can accumulate the achieved privacy [4]. An attacker monitoring multiple mix zones could, however, thwart the privacy goals of the mix network. The work most similar to ours is that of Buttyan et al. [5], who examine the relationship between the strength of the attacker and the achieved privacy, using artificial traces. Their analysis suggests that an attacker monitoring 50% of the intersections in the road network could successfully track approximately 60% of the vehicles. Our research differs from these approaches because we evaluate the upper bound on the amount of privacy provided by a single mix zone, using actual vehicle mobility data.

## 2.2 Assumptions

We assume that each VANET will use a suitable public key infrastructure where a possibly-offline but trusted certificate authority (CA) manages the identities, cryptographic keys, and certificates of all vehicles in the network. We assume that every legitimate vehicle has a unique identity, several pairs of private and public keys, and certificates corresponding to each public key.

We assume each vehicle in a VANET is equipped with a GPS receiver and can obtain the vehicle's location and the current time. Each vehicle periodically broadcasts a safety message every 100 milliseconds [15]. A safety message includes the location of the transmitting vehicle, the time of transmission, and possibly other data. Each vehicle must sign every message it sends, and distribute the corresponding certificate so that neighbors can verify the signature of each received packet [16]. We assume that each vehicle periodically changes the certificate with which it signs messages, using each certificate for a short duration and never reusing a certificate. The duration of use is beyond the scope of this paper.

We assume the location of each mix zone in a VANET is published by the trusted CA, and that both legitimate vehicles and attackers know the exact locations of the mix zones. The selection of mix zone locations is beyond the scope of this paper. We further assume that each vehicle must change its pseudonym when the vehicle passes through any mix zone. This assumption provides maximal privacy. Our analysis techniques are applicable even when some vehicles do not change pseudonyms in a mix zone.

## 2.3 Adversary Model

In this paper, we assume an attacker can monitor communications immediately adjacent to a mix zone, and as such can hear messages from entering and exiting nodes. Moreover, we assume that an attacker cannot receive safety messages from vehicles in a mix zone. Against a weak adversary, this requirement may be fulfilled using symmetric encryption [3]. Meeting this assumption against a more powerful adversary requires the use of a *silent period*, so that vehicles inside the mix zone do not transmit any safety messages at all, which may negatively impact vehicular safety within the mix zone.

We assume that an attacker can, from a safety message, identify the sending vehicle's exact lane number. This assumption is reasonable since the typical lane width in the United States is more than 3 m [17] and many off-the-shelf GPS receivers could provide accuracy within 1.5 m with 95% confidence [18].

# 3. PRIVACY METRICS

In this section, we discuss possible metrics for measuring the privacy obtained from mix zones for a given tracking algorithm. In particular, we analyze three proposed privacy metrics: entropy, success probability, and rank.

## 3.1 Entropy

Beresford et al. first used entropy to measure the user privacy provided by mix zones in location-aware services [2]. Similarly, researchers [3][4][5] have adopted entropy to measure the level of privacy provided by mix zones in VANETs. For each vehicle $v$ exiting the mix zone, the attacker calculates the probability that $v$ corresponds to an entering vehicle $i$. We denote this probability as $p_{i,v}$. The mix zone is then claimed to provide a level of privacy to vehicle $v$ defined by the entropy of $p_{i,v}$,

$$H(v) = -\sum_{i=1}^{N} p_{i,v} \log_2(p_{i,v}) \quad (1)$$

However, the correctness of the entropy value depends on the correctness of the computed probability $p_{i,v}$. Because each value of $p_{i,v}$ is computed by the tracking algorithm, and because the tracking algorithm might not correctly estimate this probability, the actual entropy obtained by a mix zone may be greater than or less than the calculated entropy.

We use the following example to illustrate our argument, as shown in Fig. 1. Let us assume vehicle X enters the mix zone from the west and vehicle Y enters the mix zone from the north simultaneously. Before vehicle X exits from the east exit, X changes its pseudonym to Y'. Similarly, vehicle Y changes its pseudonym to X' before exiting from the south exit. Let there be an attacker who tries to track each vehicle by correlating their pseudonyms before and after each vehicle crosses the mix zone.

We consider two possible attackers. Attacker *A* relies on the spatial relation between the locations where a vehicle enters and exits the mix zone to link pseudonyms. For example, the tracking algorithm of attacker *A* states that vehicles coming from the west have an equal probability of exiting from the north, south, and east sides of the mix zone. Similarly, vehicles coming from the north have an equal probability of exiting from the south, west, and east sides of the mix zone. From attacker *A*'s perspective, the probability that vehicle Y' and vehicle X are the same vehicle is 50%, the probability that vehicle Y' and vehicle Y are the same vehicle is 50% as well, and hence the entropy of vehicle Y' is 1 bit.

Attacker *B*, on the other hand, pairs the pseudonyms using a deterministic permutation. That is, attacker *B* believes that pseudonym X and X' should always belong to the same vehicle and pseudonym Y and Y' should always belong to the same vehicle. The tracking algorithm of attacker *B* may be incorrect in reality and be weaker than the tracking algorithm of attacker *A*. Nevertheless, the computed entropy of vehicle Y' is 0 since the tracking algorithm of *B* is deterministic. As a result, it seems that vehicle Y' has less privacy in the presence of attacker *B* compared to attacker *A*.

If we fix a single tracking algorithm and measure its entropy across several mix zones, the result tells us how confident the tracking algorithm is in different situations. However, as we have shown through our example, if multiple tracking algorithms are used, the resulting entropy values may provide very little insight for further analysis.

## 3.2 Success Probability and Rank

Another natural metric to measure the amount of privacy is the success probability of the attacker [5]. For each vehicle $v$ exiting the mix zone, the attacker determines the corresponding entering vehicle *i.* The attacker can make this decision by calculating the probability of the mapping of an entering vehicle $u_i$ to the exiting vehicle $v$, denoted as $p_{i,v}$, and pick the most likely mapping. The decision is then compared against ground truth. If the chosen vehicle $u_i$ is indeed the same vehicle using pseudonym $v$, then the attacker is successful. The higher the success probability is, the less privacy the mix zone offers.

A more fine-grained metric than the success probability is the rank of the actual vehicle [19]. For each exiting vehicle $v$, the tracking algorithm sorts the entering vehicles based on $p_{i,v}$. We can then compare the ground truth with the rank chosen by the tracking algorithm. For example, if the algorithm chose the actual entering vehicle as the second most likely, the algorithm would have a rank of 2 in that case. An algorithm that consistently has low rank is one that is effective at tracking vehicles through the mix zone. We can show how often a tracking algorithm is effective by computing the probability that the rank is less than some value. This metric is a generalized form of success probability: The algorithm is successful if the rank equals to 1.

Since we use two different tracking algorithms for two different mix zone configurations, we will use the rank of the actual vehicle to measure the privacy provided by mix zones.

# 4. MIX ZONES AT INTERSECTIONS

Researchers have proposed using mix zones at certain intersections. Because vehicles often simultaneously change their directions and speeds at intersections, intersections form a promising mix context [20] in which a vehicle can update its pseudonym in a way that leaks minimal information linking its two pseudonyms. In order to evaluate the privacy provided by mix zones, we first choose a tracking algorithm for an attacker and then measure the *upper bound* on the amount of privacy provided by the mix zone against that attacker.

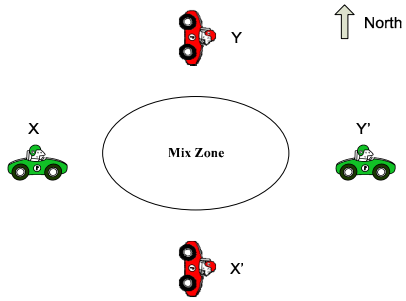


**Figure 1. Vehicles X, Y enter the mix zone and change their pseudonyms to Y' and X' respectively after exiting.**

## 4.1 Tracking at Intersections

We evaluate our algorithm using traditional machine-learning test protocols. We divide up our dataset into two phases: a training phase and a testing phase. During the training phase, an attacker (perfectly) observes incoming and outgoing traffic for a limited period of time to infer patterns with which entering vehicles correspond to exiting vehicles. During the testing phase, the attacker listens to each vehicle's location broadcasts as it enters and exits the mix zone. Then the attacker estimates the probability that a particular exiting vehicle corresponds to each of the candidate entering vehicles, using information learned during the training phase. Most proposed tracking algorithms [2][3][4][5] use two categories of information to correlate vehicles.

- Spatial correlation: The attacker observes a vehicle's location at entry and exit. Specifically, the attacker determines the probability that a node enters at location *s* and exits at location *e*.
- Temporal correlation: The attacker observes a vehicle's time of entry and exit; the elapsed time *t* follows a particular probability distribution.

We derive a more realistic traffic model by also considering lane changes and traffic lights, based upon our observations over real vehicle mobility data. As mentioned in Section 2.2, we assume that an attacker is able to identify the exact lane of traffic from which each vehicle enters and exits the mix zone. The attacker can then refine its spatial correlation technique by using lane information to help correlate entry-exit pairs. The lane-change information could provide significant advantages for the attacker. For example, based on our observations, *vehicles tend to remain in the same lane after crossing an intersection* (as long as the number of lanes remains the same before and after the intersection), and *vehicles turning right tend to stay in those lanes closer to the right edge of the road when exiting the intersection*.

Moreover, we refine the temporal correlation pattern by considering the status of traffic lights. In particular, we estimate the time that a vehicle takes to cross the intersection given the known traffic light status when the vehicle arrives at the intersection. If the light is red (we consider a yellow light to be the end of the green light cycle) when an entering vehicle wishes to cross the intersection, the vehicle has to wait a certain amount of time until the light turns green. *That vehicle would take longer to cross the intersection compared to vehicles that arrive at the intersection when the light is green.*

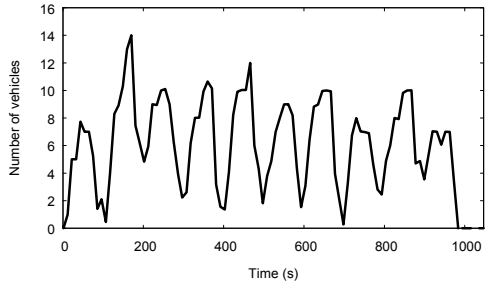


**Figure 2. The number of vehicles near the southern entry of the intersection as a function of the time.**

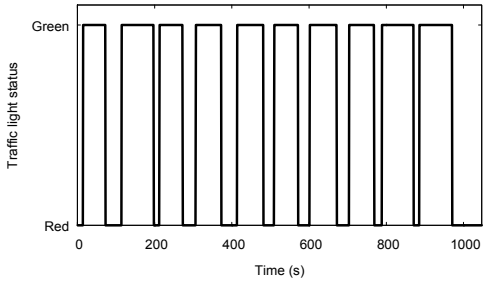


**Figure 3. The estimated light status as a function of the time.**

Our study assumes that the attacker knows, or can infer, the traffic light status. When lights are timed, an attacker can record their timing on one day and use it to estimate the status of the traffic light on another day. When lights respond in real-time to traffic conditions, the attacker can estimate the status of the light by monitoring the density and average speed at the entries and exits of the intersection. Though other intersections may use different traffic control devices such as stop signs, such intersections typically have much less traffic, making them less suitable as mix zones.

In the training phase of our tracking algorithm, we learn the probability distribution of all possible pairings of a vehicle entering the intersection in lane $i$ at entry $s$ and exiting in lane $j$ at exit $e$, denoted as $p_{si,ej}$. Next, for each vehicle arriving at the intersection, we record the traffic light status and the amount of time that the vehicle took to cross the intersection. We model the crossing time using a normal distribution, with different parameters that depend on the trajectory of the vehicle and the light status. For a vehicle arriving at the intersection at entry $s$ when the light status is $l$ and exiting at exit $e$, the probability distribution of the crossing time $q_{s,e,l}(t)$ can be denoted as

$$q_{s,e,l}(t) \sim N_l(\mu_{s,e}, \sigma_{s,e}) \qquad (2)$$

where $l \in \{Green, Red\}$.

In the testing phase of our algorithm, we represent each exit event as a 4-tuple, $(v, j, e, t_e)$ where $v$ is the pseudonym of the exiting vehicle, $j$ is the lane number, $e$ is the exit at which $v$ exits the mix zone, and $t_e$ is the time of $v$ exiting the mix zone. We represent each entering event as a 5-tuple $(f, i, s, t_s, l)$ where $f$ denotes the pseudonym of the entering vehicle, $i$ is the lane number, $s$ is the entrance at which $v$ enters the mix zone, $t_s$ is the time $v$ enters the mix zone, and $l$ is the light status at time $t_s$. For each entering and exiting event, we compute the probability that pseudonym $v$ and $f$ belong to the same vehicle as

$$P_{f,v} = p_{si,ej} q_{s,e,l}(t_e - t_s) \qquad (3)$$

Then, for each exiting vehicle $v$, we sort and rank all possible entering vehicles according to their computed probabilities.

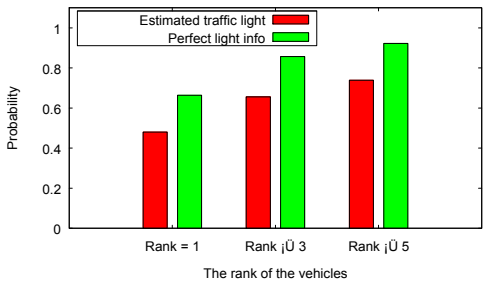


**Figure 4. The obtainable privacy provided by the mix zone created at the sample intersection of Lankershim.**

## 4.2 Case Study on Lankershim Data

In this subsection, we evaluate the level of privacy provided by a mix zone located at an intersection. We use real vehicle mobility data from the Lankershim datasets obtained from the Federal Highway Administration's (FHWA) Next Generation Simulation (NGSIM) project [21]. In the NGSIM data, both Lankershim Boulevard and Peachtree Street are multi-lane arterial roads that consist of several intersections. We choose to use the Lankershim data over the Peachtree data, because the Peachtree data has very little cross traffic in the east-west direction and all of the Peachtree intersections are either controlled by timer based traffic lights or have no traffic lights at all.

In the Lankershim data, we choose the intersection of Universal Hollywood Drive and Lankershim Boulevard because this intersection has the densest traffic. The traffic at this intersection is coordinated by adaptive traffic lights, which dynamically changes status based on the traffic condition around the intersection. The Lankershim data includes two sets of data, each lasting about 15 minutes. Across both data sets, 2,230 vehicles cross this intersection. We use the data starting from 8:30 am as the training dataset and the trace starting from 8:45 am as the testing dataset. We define a circular mix zone centered at the middle of the intersection. The radius of the mix zone is 24 meters, which covers approximately the entire intersection.

During the training phase, we assume the attacker knows the precise traffic light status at any time, which we derived from the recorded Lankershim video data. In the testing phase, we *estimate* the traffic light status every second during the entire 15-minute period. In our evaluation, we only use the change in vehicle density to estimate the light status. We monitor the number of vehicles near the southern entry of the intersection every second, as depicted in Fig. 2. We derive the status of the traffic light controlling the northbound traffic at every second, as depicted in Fig. 3 (details are omitted here due to space constraints). For comparison, we let the attacker know the exact light status at any second and run the same experiment again.

We plot the fraction of cases where the actual vehicle's rank was 1, at most 3, and at most 5. Fig. 4 shows our results. The attacker is able to correctly identify the entering vehicle for about 48.0% of the exiting vehicles. We also find that the probability of success could be significantly improved if the attacker can learn the exact times at which the traffic light changes. Simulations in previous work suggest if there are on average 7 vehicles crossing the intersection under high traffic congestion scenario, then less than 30% of the vehicles could be tracked [4]. However, our evaluation based on actual vehicle mobility data shows that a mix zone at a busy intersection only offers limited privacy. Vehicles crossing the mix zone would obtain even less privacy if the attacker could get accurate traffic light status along both directions of the intersection, or if the attacker uses more sophisticated tracking algorithms.

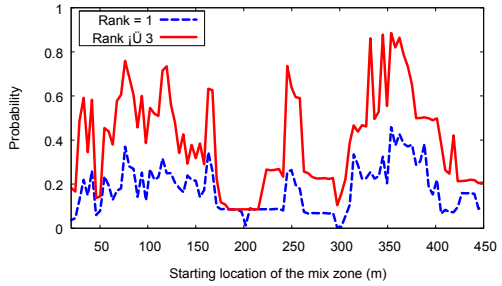


**Figure 5. The obtainable privacy as a function of the starting location of the mix zone ($l$ = 50 m, lane 5).**

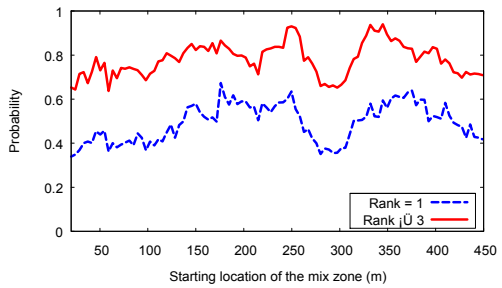


**Figure 6. The obtainable privacy as a function of the starting location of the mix zone when 5% of the entering vehicles have already been tracked ($l$ = 50 m, lane 5).**

# 5. MIX ZONES ON STRAIGHT ROADS

Unlike at intersections, vehicles on straight roads rarely change directions. However, if the vehicle density on the roadway is relatively high, drivers might change lanes and speed, making these locations possible candidates for mix zones. In this section, we evaluate the effectiveness of mix zones created on a high density straight highway using the I-80 data from NGSIM.

## 5.1 Tracking on Straight Roads

In order to evaluate the effectiveness of mix zones on straight roads, we choose a tracking algorithm that correlates vehicles based on the sequence of vehicles entering and exiting the road section, similar to the idea proposed in [22].

Because straight-road tracking depends on typical driver behavior, errors can accumulate without the periodic injection of ground truth. An attacker can take advantage of easily tracked vehicles in order to obtain this ground truth. Some public vehicles, such as police cars, might be tracked easily if they send out distinct messages. Other VANET applications may also disclose information sufficient for such tracking. Other RF devices in a car, such as electronic toll collection or an RF tire pressure monitor, could also be used for tracking. These easily tracked vehicles can help an attacker reduce accumulations of error, as we show in our evaluations.

Our tracking algorithm is as follows. For each lane $k$ on a straight road, we maintain a counter $CI_k$ and assign a unique sequence number for every vehicle entering lane $k$. The counter $CI_k$ starts from zero and is incremented by one when a vehicle enters the mix zone in lane $k$. Similarly, we maintain another counter $CO_k$ and assign a unique sequence number for every vehicle exiting the mix zone in lane $k$. The counter $CO_k$ starts from zero and is incremented by one when a vehicle exits the mix zone in lane $k$. The attacker could initiate tracking when there are no vehicles on the road, e.g., at some time during the night. The attacker then tries to track a vehicle by pairing the exit sequence number with the same entry sequence number; that is, the *n-th* vehicle entering the mix zone in lane $k$ is paired with the *n-th* vehicle exiting the mix zone in lane $k$. One disadvantage of this approach is that if vehicle $v$ enters the mix zone in lane $k$ and later changes to lane $q$, all the vehicles following $v$ in lane $k$ will have their exit sequence number offset by 1. These errors could accumulate if additional vehicles in lane $k$ change lanes. However, as previously mentioned, some vehicles could be tracked using other approaches; these vehicles could be used to remove the accumulated errors. Whenever a trackable vehicle enters and exits the mix zone in lane $k$, we reset the entry counter $CI_k$ to zero when the vehicle enters the mix zone and reset the exit counter $CO_k$ to zero when the vehicle exits the mix zone.

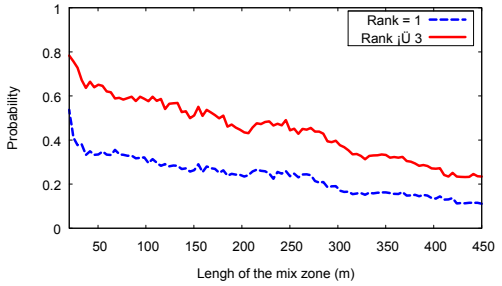


**Figure 7. The obtained privacy as a function of the length of the mix zone (lane 5).**

## 5.2 Case Study on I-80 Data

In this subsection, we study the privacy obtainable from mix zones built on a high-density straight road. We assume that the attacker uses the tracking algorithm described above; thus our evaluation results serve as an upper bound of obtainable privacy. In the NGSIM data, both I-80 and US-101 are major multi-lane freeways. We choose the I-80 data over the US-101 data because the I-80 data has denser traffic.

The I-80 data represents a stretch of roadway approximately 503 meters in length, including an on-ramp at Powell Street. There are six through lanes on I-80, excluding the on-ramp. Lane numbering starts from the left-most lane, which we number lane 1. Lane 6 is the closest to the on-ramp. In the I-80 data, we choose to study the 5:00 pm – 5:15 pm dataset because it contains the densest traffic among all three of the datasets, consisting of 1836 vehicles. The average density of the trace is 52.1 vehicles per km per lane. We exclude the traffic coming from the on-ramp (205 vehicles) from our evaluation.

We define a rectangular area on I-80 as the mix zone; the rectangular area is $l$ meters in length, consisting of all six straight lanes. Both the starting location of the mix zone on I-80 and the value of $l$ vary. We choose to evaluate the privacy of the vehicles entering the mix zone in lane 5, because of all six lanes, lane changes take place most frequently in lane 5. Fig. 5 shows the privacy obtained from mix zones built on I-80, where $l$ = 50 m. We build the first mix zone starting from the southern edge of the I-80 data set and increment the starting location of the mix zone by 3 meters (approximately the length of one vehicle) for every run. The results suggest that the privacy obtained from the mix zone is highly dependent on the location of the mix zone.

We then investigate the impact of easily tracked vehicles on the privacy of other nodes around them. We uniformly at random choose 5% of the vehicles entering the mix zone in lane 5 and assume they can be easily tracked. We performed 100 runs for each mix zone and averaged the results. Comparing Fig. 6 to Fig. 5 shows that the attacker is much more successful if some vehicles can be easily tracked. Over 60% of all vehicles exiting the mix zone in lane 5 are one of the three highest ranked entering vehicles. For comparison, we also carry out the same evaluation in lane 1 and lane 2 (detailed results omitted due to space constraints). Vehicles in

these lanes could not obtain as much privacy as those in lane 5, because lane 1 and lane 2 are farther away from the on-ramp than lane 5 and hence vehicles in lane 1 and 2 are less affected by the traffic coming from the on-ramp. In other words, true straight-line roads offer little privacy.

Finally we evaluate the impact the size of the mix zone has on the obtainable privacy. We fix the starting position of the mix zone near the southern edge of the I-80 dataset and gradually increase the length of the mix zone from 20 m to 500 m. As we expected, Figure 7 suggests that larger mix zones provide better privacy.

# 6. CONCLUSION

We took realistic traffic models into consideration and evaluated the effectiveness of mix zones at intersections and on straight roads using actual mobility traces. Contrary to implications of prior work, our results suggest that a single mix zone provides limited privacy because of the inherent lack of randomness in vehicle mobility. However, a vehicle might obtain more privacy by consecutively traversing several mix zones. Evaluating the effectiveness of a network of mix zones is beyond the scope of this paper and we leave this to future work. We believe that deciding whether or not the privacy provided by a single mix zone is sufficient should also be a policy question, which depends on the opinion of VANET policy-makers.

# 7. References


1. **The CAMP Vehicle Safety Communications Consortium.** *Vehicle Safety Communications Project Final Report: Identify Intelligent Vehicle Safety Applications Enabled by DSRC.* s.l. : U.S. Department of Transportation, National Highway Traffic Safety Administration, 2005. DOT HS 809 859.

2. *Mix-zones: User privacy in location-aware services.* **Beresford, A. R. and Stajano, F.** s.l. : IEEE, 2004. Proceedings of the First IEEE International Workshop on Pervasive Computing and Com-munication Security (PerSec).

3. *Mix-Zones for Location Privacy in Vehicular Networks.* **Freudiger, J., et al.** 2007. Proceedings of WiN-ITS.

4. *On the Optimal Placement of Mix Zones.* **Freudiger, J., Shokri, R. and Hubaux, J.-P.** 2009. Proceedings of the 9th Privacy Enhancing Technologies Symposium (PETS).

5. *On the effectiveness of changing pseudonyms to provide location privacy in VANETs.* **Buttyan, L., Holczer, T. and Vajda, I.** 2007. Proceedings of the 4th European Workshop on Security and Privacy in Ad Hoc and Sensor Networks (ESAS).

6. *Privacy issues in vehicular ad hoc networks.* **Dötzer, F.** 2005. Proceedings of the Workshop on Privacy Enhancing Technologies (PET).

7. **Hubaux, J.-P., Capkun, S. and Luo, J.** The Security and Privacy of Smart Vehicles. *IEEE Security & Privacy Magazine.* 2004, Vol. 2, 3.

8. *The Security of Vehicular Ad Hoc Networks.* **Raya, M. and Hubaux, J.-P.** 2005. Proceedings of the Third ACM Workshop on the Security of Ad Hoc and Sensor Networks (SASN).

9. *Architecture for Secure and Private Vehicular Communications.* **Papadimitratos, P., et al.** 2007. Proceedings of the International Conference on ITS Telecommunications (ITST).

10. *Enhancing wireless location privacy using silent period.* **Huang, L., et al.** 2005. Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC).

11. *Caravan: Providing location privacy for VANET.* **Sampigethaya, K., et al.** 2005. Proceedings of the 3rd Annual Conference on Embedded Security in Cars (ESCAR).

12. *Amoeba: Robust location privacy scheme for VANET.* **Sampigethaya, K., et al.** 8, 2007, IEEE Journal on Selected Areas in Communications, Vol. 25, pp. 1569--1589.

13. *Swing and Swap: User-centric approaches towards maximizing location privacy.* **Li, M., et al.** 2006. Proceedings of the 5th ACM workshop on Privacy in electronic society (WPES).

14. *Location privacy in pervasive computing.* **Beresford, A. R. and Stajano, F.** 1, 2003, IEEE Pervasive Computing, Vol. 3, pp. 46-55.

15. **IEEE.** *5.9 GHz DSRC: Dedicated Short-Range Communications.* http://grouper.ieee.org/groups/scc32/dsrc/.

16. *IEEE 1609.2-Standard for Wireless Access in Vehicular Environments (WAVE) - Security Services for Applications and Management Messages.*

17. **Stein, William J. and Neuman, Timothy R.** *Mitigation Strategies for Design Exceptions.* http://safety.fhwa.dot.gov/geometric/pubs/mitigationstrategies/fhwa_sa_07011.pdf.

18. *Wide-Area Augmentation System Performance Analysis Report, Reporting Period: January 1 to March 31.* 2009. http://www.nstb.tc.faa.gov/.

19. *Location Privacy in Wireless Networks.* **Hu, Y.-C. and Wang, H. J.** 2005. Proceedings of the ACM SIGCOMM Asia Workshop.

20. *Privacy in VANETs using changing pseudonyms - ideal and real.* **Gerlach, M. and Gäuttler, F.** 2007. Proceedings of the 65th Vehicular Technology Conference (VTC).

21. NGSIM. Next Generation Simulation. [Online] http://www.ngsim.fhwa.dot.gov/.

22. *A Method for Deter-mining Real-Time Travel Times on Motorways.* **Westerman, M. and Immers, L.** 1992. Proceedings of the 25th International Symposium on Automotive Technology and Automation (ISATA). pp. 221--228.