



# Location Privacy in Vehicular Ad Hoc Networks

N.K.Prema,  
Research Scholar/CSE  
PRIST University,  
Thanjavur, Tamil Nadu

Dr.A.Arul Lawrence S.K  
Professor /CSE  
Rajiv Gandhi Institute of Technology  
Bangalore -32

---

**Abstract-** *One approach to solve this problem is that the vehicles broadcast their messages under pseudonyms that they change with some frequency. The change of a pseudonym means that the vehicle changes all of its physical and logical addresses at the same time. Indeed, in most of the applications, the important thing is to let other vehicles know that there is a vehicle at a given position moving with a given speed, but it is not really important which particular vehicle it is. Thus, using pseudonyms is just as good as using real identifiers as far as the functionality of the applications is concerned. Obviously, these pseudonyms must be generated in such a way that a new pseudonym cannot be directly linked to previously used pseudonyms of the same vehicle. Unfortunately, only changing pseudonyms are largely ineffective against a global eavesdropper that can hear all communications in the network.*

**Keywords-** *AdHoc Networks, eavesdropping, networks*

---

## 1 INTRODUCTION

In this Paper, I investigate what level of privacy a driver can achieve in Vehicular Ad Hoc Networks (VANET). More specifically, in the first half of this Paper, I investigate how can a local eavesdropping attacker trace the vehicles based on their frequently sent status information. In the second half of this Paper (from Section 3.4), I go a little further in terms of strength of attacker, and check what can a global eavesdropping attacker do. After realizing its broad capabilities, I suggest an algorithm, which can greatly reduce the attackers success rate. Recently, initiatives to create safer and more efficient driving conditions have begun to draw strong support in Europe [COM], in the US [VSC], and in Japan [ASV]. Vehicular communications will play a central role in this effort, enabling a variety of applications for safety, traffic efficiency, driver assistance, and entertainment. However, besides the expected benefits, vehicular communications also have some potential drawbacks. In particular, many envisioned safety related applications require that the vehicles continuously broadcast their current position and speed in so called *heart beat* messages. This allows the vehicles to predict the movement of other nearby vehicles and to warn the drivers if a hazardous situation is about to occur. While this can certainly be advantageous, an undesirable side effect is that it makes it easier to track the physical location of the vehicles just by eavesdropping these heart beat messages.

Such an adversary can predict the movement of the vehicles based on the position and speed information in the heart beat messages, and use this prediction to link different pseudonyms of the same vehicle together with high probability. For instance, if at time  $t$ , a given vehicle is at position  $\vec{p}$  and moves with speed  $\vec{v}$ , then after some short time  $\tau$ , this vehicle will most probably be at position  $\vec{p} + \tau \cdot \vec{v}$ . Therefore, the adversary will know that the vehicle that reports itself at (or near to) position  $\vec{p} + \tau \cdot \vec{v}$  at time  $t + \tau$  is the same vehicle as the one that reported itself at position  $\vec{p}$  at time  $t$ , even if in the meantime, the vehicle changed pseudonym.

This problem can be solved with some silent periods. This is discussed in the second part of this Paper on the other hand, the assumption that the adversary can eavesdrop all communications in the network is a very strong one. In many situations, it is more reasonable to assume that the adversary can monitor the communications only at a limited number of places and only in a limited range. In this case, if a vehicle changes its pseudonym within the non-monitored area, then there is a chance that the adversary loses its trace. My goal in the first half of the Paper is to characterize this chance as a function of the strength of the adversary (i.e., its monitoring capabilities). In the second part of the Paper, I assume a relatively small area, where a global eavesdropping is reasonable. I analyze what a global attacker can do, and suggest a simple algorithm to reduce the capabilities of a global attacker. In particular, my main contributions are the following: I define a model in which the effectiveness of changing pseudonyms can be studied. I emphasize that while changing pseudonyms has already been proposed in the literature as a countermeasure to track vehicles to the best of my knowledge, the effectiveness of this method has never been investigated rigorously in this context. My model is based on the concept of the *mix zone*. This concept was first introduced in, but again, to the best of my knowledge, it has not been used in the context of vehicular networks so far. I characterize the tracking strategy of the adversary in the mix zone model, and I introduce a metric to quantify the level of privacy provided by the mix zone. I report on the results of an extensive simulation where I used my model to determine the level of privacy achieved in realistic scenarios. In particular, in my simulation, I used a rather complex road map, generated traffic with realistic parameters, and varied the strength of the adversary by varying the number of her monitoring points.

As expected, my simulation results confirm that the level of privacy decreases as the strength of the adversary increases. However, in addition to this, my simulation results provide detailed information about the relationship between the strength of the adversary and the level of privacy achieved by changing pseudonyms. I provide a breakdown of the requirements that a system must address in order to provide privacy. The aim is to provide an analytical framework that future researchers can use to concisely state which aspects of privacy a new proposal does or does not address. In Section 3.6, I propose an approach for implementing mix zones that does neither require extensive RSU support nor complex communication between vehicles, and that does not endanger safety-of-life to any significant extent, while providing both *syntactic mixing* and *semantic mixing* (in the language of Section 3.5). To my knowledge, this is the first proposal that provides for semantic mixing while at the same time addressing the safety-of-life concerns that naturally arise when a vehicle tries to obscure its path. The key insights are simply that vehicles traveling at a low speed are less likely to cause fatal accidents, and that vehicles will be traveling at a low speed at natural mix-points such as signalled intersections. The main body of experimental work in Section 3.6 is therefore an investigation of the consequences for the untraceability of vehicles if they stop sending heartbeat messages when their speed drops below a certain threshold and change all their identifiers after such silent periods. I call my scheme SLOW, which stands for silence at low speeds. (I note that of course SLOW is not a full solution to untraceability, as it does not cover the safe use of silent periods at high speeds; other techniques will need to be used to give untraceability in this case). The organization of the Paper is the following: in Section 3.2, I introduce the mix zone model, I define the behavior of the adversary in this model, and I introduce my privacy metric. In Section 3.3, I describe my simulation setting and the simulation results for the mix zones. In Section 3.4 I introduce the global attacker scenario. Then I introducing my overall analytical framework in Section 3.5. Next, in Section 3.6, I introduce my attacker model and my proposed solution, and in Section 3.7, I present the results of my experiments showing that my approach does indeed make tracing of vehicles hard for the attacker, and that it is usable in the real world. Finally, I report on some related work in Section 3.8, and conclude the Paper in Section 3.9.

## 2 MODEL OF LOCAL ATTACKER AND MIX ZONE

### 2.1 THE CONCEPT OF THE MIX ZONE

I consider a continuous part of a road network, such as a whole city or a district of a city. I assume that the adversary installed some radio receivers at certain points of the road network with which she can eavesdrop the communications of the vehicles, including their heart beat messages, in a limited range. On the other hand, outside the range of her radio receivers, the adversary cannot hear the communications of the vehicles. Thus, the road network is divided into two distinct regions: the observed zone and the unobserved zone. Physically, these zones may be scattered, possibly consisting of many observing *spots* and a large unobserved area, but logically, the scattered observing spots can be considered together as a single observed zone. This is illustrated on the left hand side of Figure 3.1.

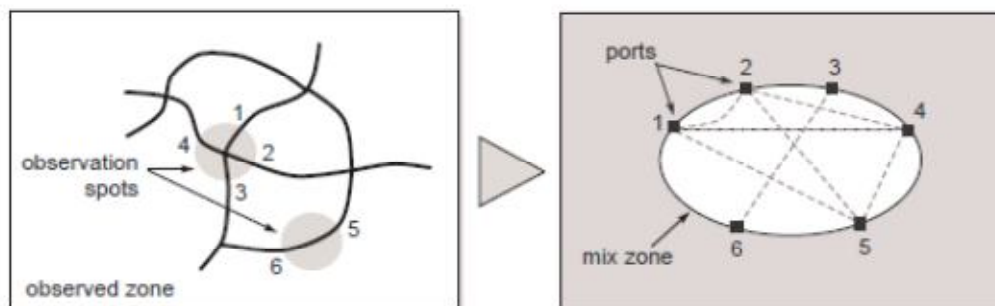


Figure 3.1: **On the left hand side:** The figure illustrates how a road network is divided into an observed and an unobserved zone in the model. In the figure, the observed zone is grey, and the unobserved zone is white. The unobserved zone functions as a *mix zone*, because the vehicles change pseudonyms and mix within this zone making it difficult for the adversary to track them. **On the right hand side:** The figure illustrates how the road network on the left can be abstracted as single mix zone with six ports.

Note that the vehicles do not know where the adversary installed her radio receivers, or in other words, when they are in the observed zone. For this reason, we can assume that the vehicles continuously change their pseudonyms. In this part of the Paper, we can abstract away the frequency of the pseudonym changes, and we can simply assume that it is high enough so that every vehicle surely changes pseudonym while in the unobserved zone. I intend to relax this assumption in my future work. Since the vehicles change pseudonyms while in the unobserved zone, that zone functions as a *mix zone* for vehicles (see the right hand side of Figure 3.1 for illustration). A mix zone [Beresford and Stajano, 2003; Beresford and Stajano, 2004] is similar to a mix node of a mix network [Chaum, 1981], which changes the encoding and the order of messages in order to make it difficult for the adversary to link message senders and message receivers. In my case, the mix zone makes it difficult for the adversary to link the vehicles that emerge from the mix zone to those that entered it earlier.

Thus, the mix zones makes it difficult to track vehicles. On the other hand, based on the observation that I made in the Introduction, I assume that the adversary can track the physical location of the vehicles while they are in observed zone, despite the fact that they may change pseudonyms in that zone too. Since the vehicles move on roads, they cannot cross the border between the mix zone and the observed zone at any arbitrary point. Instead, the vehicles cross the border where the roads cross it. We can model this by assuming that the mix zone has *ports*, and the vehicles can enter and exit the mix zone only via these ports. For instance, on the right hand side of Figure 3.1, the ports are numbered from 1 to 6.

## 2.2 THE MODEL OF THE MIX ZONE

While the adversary cannot observe the vehicles within the mix zone, we can assume that she still has some knowledge about the mix zone. This knowledge is subsumed in a model that consists of a matrix  $Q = [q_{ij}]$  of size  $M \times M$ , where  $M$  is the number of ports of the mix zone, and  $M^2$  discrete probability density functions  $f_{ij}(t)$  ( $1 \leq i, j \leq M$ ).  $q_{ij}$  is the conditional probability of exiting the mix zone at port  $j$  given that the entry point was port  $i$ .  $f_{ij}(t)$  describes the probability distribution of the delay when traversing the mix zone between port  $i$  and port  $j$ . We can assume that time is slotted, that is why  $f_{ij}(t)$  is a discrete function. I note here, that it is unlikely for an attacker to achieve such a comprehensive knowledge of the mix zone. However it is not impossible with Comprehensive real world measurements to approximate the needed probabilities and functions. In the rest of the Paper, we can consider the worst case (as it is advisable in the field of security), the attacker knows the model of the mix zone.

## 2.3 THE OPERATION OF THE ADVERSARY

The adversary knows the model of the mix zone and she observes *events*, where an event is a pair consisting of a port (port number) and a time stamp (time slot number). There are entering events and exiting events corresponding to vehicles entering and exiting the mix zone, respectively. Naturally, an entering event consists of the port where the vehicle entered the mix zone, and the time when this happened. Similarly, an exiting event consists of the port where the vehicle left the mix zone, and the time when this happened. The general objective of the adversary is to relate exiting events to entering events. More specifically, in the model, the adversary picks a vehicle  $v$  in the observed zone and tracks its movement until it enters the mix zone. In the following, I denote the port at which  $v$  entered the mix zone by  $s$ . Then, the adversary observes the exiting events for a time  $T$  such that the probability that  $v$  leaves the mix zone before  $T$  is close to 1 (i.e.,  $\Pr\{tout < T\} = 1 - \epsilon$ , where  $\epsilon$  is a small number, typically, in the range of 0.005–0.01, and  $tout$  is the random variable denoting the time at which the selected vehicle  $v$  exits the mix zone). For each exiting vehicle  $v'$ , the adversary determines the probability that  $v'$  is the same as  $v$ . For this purpose, she uses her observations and the model of the mix zone. Finally, she decides which exiting vehicle corresponds to the selected vehicle  $v$ . The decision algorithm used by the adversary is intuitive and straightforward: the adversary knows that the selected vehicle  $v$  entered the mix zone at port  $s$  and in timeslot 0. For each exiting event  $k = (j, t)$  that the adversary observes afterwards, she can compute the probability  $p_{jt}$  that  $k$  corresponds to the selected vehicle as  $p_{jt} = q_{sj}f_{sj}(t)$  (i.e., the probability that  $v$  chooses port  $j$  as its exit port given that it entered the mix zone at port  $s$  multiplied by the probability that it covers the distance between ports  $s$  and  $j$  in time  $t$ ). The adversary decides for the vehicle for which  $p_{jt}$  is maximal. The adversary is successful if the decided vehicle is indeed  $v$ . Indeed, the above described decision algorithm realized the Bayesian decision (see the Section 3.2.4 for more details). The importance of this fact is that the Bayesian decision minimizes the error probability, thus, it is in some sense the ideal decision algorithm for the adversary.

## 2.4 ANALYSIS OF THE ADVERSARY

In this section, I show that the decision algorithm of the adversary described in Subsection 3.2.3 realizes a Bayesian decision. The following notations are used:

$k$  is an index of a vector. Every port-timeslot pair can be mapped to such an index and  $k$  can be mapped back to a port-timeslot pair. Therefore indices and port-timeslot pairs are interchangeable, and in the following discussion, I always use the one which makes the presentation simpler.

- $k \in 1 \dots M \cdot T$ , where  $M$  is the number of ports, and  $T$  is the length of the attack measured in timeslots.
- $C = [ck]$  is a vector, where  $ck$  is the number of cars leaving the mix zone at  $k$  during the attack.
- $N$  is the number of cars leaving the mix zone before timeslot  $T$  (i.e.,  $N = \sum_{k=1}^{MT} ck$ ).
- $ps(k)$  is the probability of the event that the target vehicle leaves the mix zone at  $k$  (port and time) conditioned on the event that it enters the zone at port  $s$  at time 0. The attacker exactly knows which port is  $s$ . Probability  $ps(k)$  can be computed as:  $ps(k) = q_{sj}f_{sj}(t)$ , where port  $j$  and timeslot  $t$  correspond to index  $k$ .
- $p(k)$  is the probability of the event that a vehicle leaves the mix zone at  $k$  (port and time). This distribution can be calculated from the input distribution and the transition probabilities:
- $p(k) = \sum_{s=1}^M ps(k)$ .
- $\Pr(k/C)$  is the conditional probability that the target vehicle left the mix zone at time and port defined by  $k$ , given that the attacker's observation is vector  $C$ .

We must determine for which  $k$  probability  $\Pr(k/C)$  is maximal. Let us denote this  $k$  with  $k^*$ . The probability  $\Pr(k/C)$  can be rewritten, using the Bayes rule:  $\Pr(k/C) = \Pr(C/k)ps(k) / \Pr(C)$   
 Then  $k^*$  can be computed as:  $k^* = \arg \max_k \Pr(C/k)ps(k) / \Pr(C) = \arg \max_k \Pr(C/k)ps(k)$

$K \Pr(C/k)ps(k)$

$\Pr(C/k)$  has a multinomial distribution with a condition that at least one vehicle (the target of the attacker) must leave the mix zone at  $k$ :

$$\Pr(C/k) = \frac{N!}{c_1! \dots c_k-1!(c_k-1)!c_{k+1}! \dots c_M T!} p(k)^{c_k-1} \prod_{j=1, j \neq k}^M p(j)^{c_j}$$

$\Pr(C/k)$  can be multiplied and divided by  $p(k)^{c_k}$  to simplify the equation:  
 $\Pr(C/k) = \frac{c_k p(k)^{c_k}}{N! c_1! \dots c_M T!} \prod_{j=1}^M p(j)^{c_j}$

where the bracketed part is a constant, which does not have any effect on the maximization, thus it can be omitted.  $K^* = \arg \max_k c_k p(k)^{c_k} = \arg \max_k c_k p(k)^{c_k} N ps(k) = \arg \max_k c_k p(k)^{c_k} ps(k)$  where  $b p(k)$  is the empirical distribution of  $k$  (i.e.,  $b p(k) = c_k/N$ ). If the number of vehicles in the mix zone is large enough, then  $c_k p(k)^{c_k} \approx 1$ . Thus correctness of the intuitive algorithm described in Subsection 3.2.3 holds:  $k^* = \arg \max_k ps(k)$  This means that if many vehicles are traveling in the mix zone, then the attacker must choose the vehicle with the highest  $ps(k)$  probability.

### 2.5 THE LEVEL OF PRIVACY PROVIDED BY THE MIX ZONE

There are various metrics to quantify the level of privacy provided by the mix zone (and the fact that the vehicles continuously change pseudonyms). A natural metric in the model is the success probability of the adversary when making her decision as described above. If the success probability is large, then the mix zone and changing pseudonyms are ineffective. On the other hand, if the success probability of the adversary is small, then tracking is difficult and the system ensures location privacy.

We can note that the level of privacy is often measured using the anonymity set size as the metric however, in this case, this approach cannot be used. The problem is that as described above, with probability  $\epsilon$ , the selected vehicle  $v$  is not in the set  $V$  of vehicles exiting the mix zone during the experiment of the adversary, and therefore, by definition,  $V$  cannot be the anonymity set for  $v$ . Although, the size of  $V$  could be used as a lower bound on the real anonymity set size, there is another problem with the anonymity set size as privacy metric. Namely, it is an appropriate privacy metric only if each member of the set is equally likely to be the target of the observation, however, as we will see in Section 3.3, this is not the case in my model. Obviously, the success probability of the adversary is very difficult to determine analytically due to the complexity of the model. Therefore, I ran simulations to determine its empirical value in realistic situations. The simulation setting and parameters, as well as the simulation results are described in the next section.

## 3 SIMULATION OF MIX ZONE

The purpose of the simulation is to get an estimation of the success probability of the attacker in realistic scenarios. In this section, I first describe the simulation settings, and then, I present the simulation results.

### 3.1 SIMULATION SETTINGS

The simulation was carried out in three main phases. In the first phase, I generated a realistic map, where the vehicles moved during the simulation. This map was generated by MOVE [Karnadi *et al.*, 2005], a tool that allows the user to quickly generate realistic mobility models for vehicular network simulations. My map is illustrated in Figure 3.2. In fact, it is a simplified map of Budapest, the capital of Hungary, and it contains the main roads of the city. I believe that despite of the simplifications, this map is still complex enough to get realistic traffic scenarios. The second phase of the simulation was to generate the movement of the vehicles on the generated map. This was done by SUMO [Krajzewicz *et al.*, 2002], which is an open source microtraffic simulator, developed by the Center for Applied Informatics (ZAIK) and the Institute of Transport Research at the German Aerospace Center. SUMO dumps the state of the simulation in every time step into files. This state dump contains the location and the velocity of every vehicle during the simulation. In the third phase of the simulation, I processed the state dump generated by SUMO, and simulated the adversary. This part of the simulation was written in Perl, because Perl scripts can easily process the XML files generated by SUMO. Note that for the purpose of repeatability, I implemented the adversary as follows. First, I defined the observation spots (position and radius) of the adversary in a configuration file. Then, I let the adversary build her model of the mix zone (i.e., the complement of its observation spots) by allowing her to track the vehicles as if they do not change their pseudonyms. In effect, the adversary's knowledge is represented by a set of two dimensional tables. Each table  $K(i)$  corresponds to a port  $i$  of the mix zone, and contains empirical probabilities. More specifically, the entry  $K(i)_{jt}$  of table  $K(i)$  contains the empirical probability that a vehicle exits the mix zone at port  $j$  in time  $t$  given that it entered the mix zone at port  $i$  at time 0. The size of the tables is  $M \times T$ , where  $M$  is the number of the ports of the mix zone and  $T$  is the duration of the learning procedure defined as the time until which every observed vehicle left the mix zone.

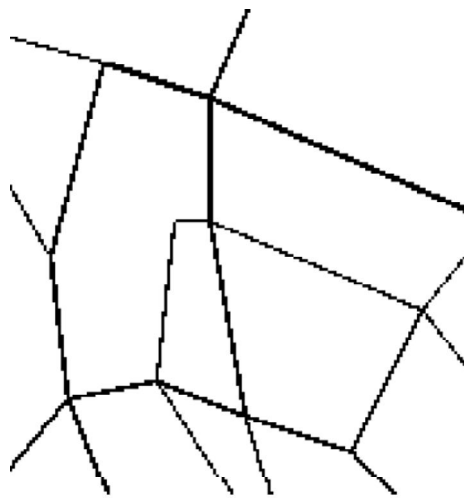


Figure 3.2: Simplified map of Budapest generated for the simulation

Once the adversary's knowledge is built, she could use that for making decisions as described above in Section 3.2. I executed several simulation runs in order to get an estimation for the success probability of the adversary. Experiments with adversaries of different strength are made, where the strength of the adversary depends on the number of her eavesdropping receivers. In the simulations, all receivers were deployed in the middle of the junctions of the roads. The eavesdropping radius of the receivers was set to 50 meter. The number of the receivers varied between 5 and 59 with a step size of 5 (note that the map contains 59 junctions). Always the junctions with the highest traffic was chosen as the observation spots of the adversary (for instance, when the adversary had ten receivers, I chose the first ten junctions with the largest traffic). In addition to the strength of the adversary, the intensity of the traffic is varied. More specifically, I simulated three types of traffic: low, medium, and high. Low traffic means that in each time step 250 vehicles are emitted into the traffic flow, medium traffic is defined as 500 vehicles are emitted into the flow, and in case of high traffic 750 vehicles are emitted. For each simulation setting (strength of the adversary and intensity of the road traffic) 100 simulations were performed.

### 3.2 SIMULATION RESULTS

Figure 3.3 contains the resulting success probabilities of the adversary as a function of her strength. The different curves belong to different traffic intensities. The results are quite intuitive: we can conclude that the stronger the adversary, the higher her success probability. Note, however, that from above a given strength, the success probability saturates at about 60 %. Higher success probabilities cannot be achieved, because the order of the vehicles may change between junctions without the adversary being capable of tracking that. Note also that the saturation point is reached with the control of only the half of the junctions. The intensity of the traffic is much less important parameter, than the strength of the attacker. The success probability of the attacker is nearly independent from the intensity of the traffic above a given attacker strength.

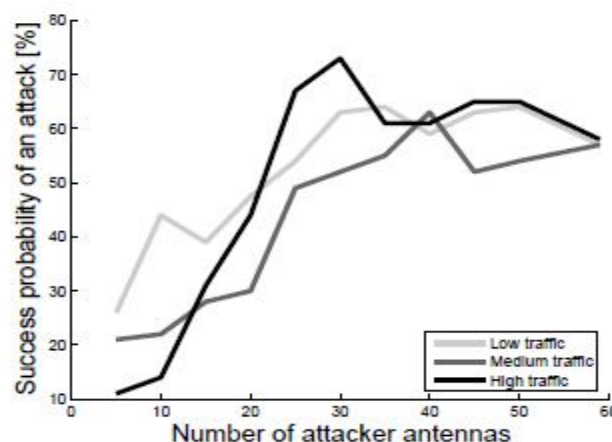


Figure 3.3: Success probabilities of the adversary as a function of her strength. The three curves represent three different scenarios (the darker the line, the more intensive the traffic).

The dark bars in Figure 3.4 show how the size of the set  $V$  of the vehicles that exit the mix zone during the observation period and from which the adversary has to decide to the selected vehicle varies with the strength of the adversary. The three sub-figures are related to the three different traffic situations (low traffic – left, medium traffic – middle, high traffic – right). While the size of  $V$  seems to be large (which seemingly makes the adversary’s decision difficult), it is also interesting to examine how uniform this set  $V$  is in terms of the probabilities assigned to the vehicles in  $V$ . Recall that the adversary computes a probability  $p_{jt}$  for each vehicle  $v'$  in  $V$ , which is the probability of  $v' = v$ . These probabilities can be normalized to obtain a distribution, and the entropy of this distribution can be computed. From this entropy, I computed the effective size of  $V$  (i.e., the size to which  $V$  can be compressed due to the non-uniformity of the distribution over its members), and the light bars in the figure illustrate the obtained values. As we can see, the effective size of  $V$  is much smaller than its real size, which means that the distribution corresponding to the members of  $V$  is highly non-uniform. This is the reason why the adversary can be successful.

### 3.3 GLOBAL ATTACKER

In the following part of this Paper, I assume a global eavesdropping attacker instead of a local attacker. A global eavesdropping attacker can hear all of the messages sent by the vehicles. This is a more challenging task, compared to the local attacker scenario. However, requires the use of significant infrastructure. By replacing [Freudiger *et al.*, 2007]’s cryptographic mix zones with zones of silence I address semantic mixing and infrastructure requirements simultaneously. In the following, in Section 3.5, I give a framework, where the minimal requirements for providing privacy for vehicles is analyzed. I present the results of my experiments showing that my approach does indeed make tracing of vehicles hard for the attacker, and that it is usable in the real world.

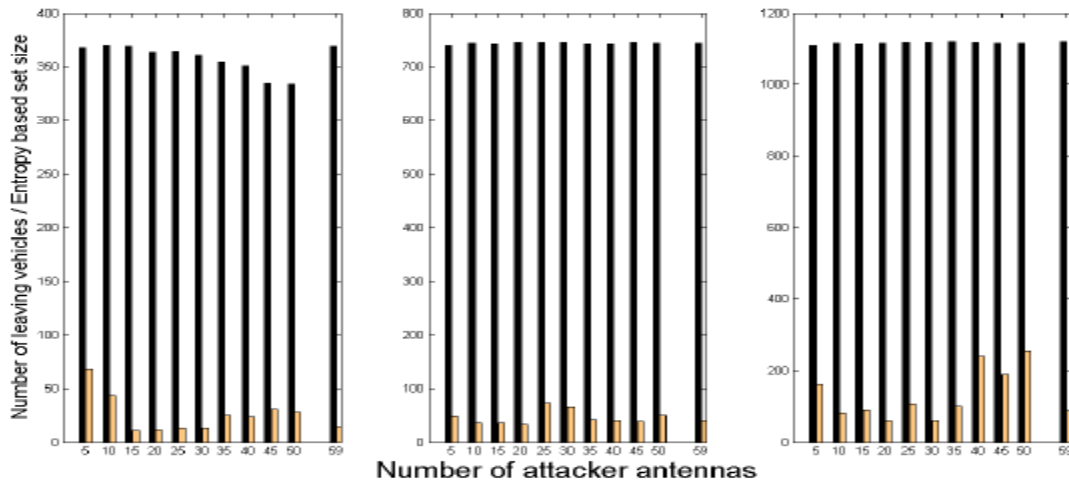


Figure 3.4: The dark bars show how the size of the set  $V$  of the vehicles that exit the mix zone during the observation period varies with the strength of the adversary (y axis: number of attacker antennas). The three sub-figures are related to the three different traffic situations (low traffic – left, medium traffic – middle, high traffic – right). The light bars illustrate the effective size of  $V$ . As we can see, the effective size is much smaller than the real size, which means that distribution corresponding to the members of  $V$  is highly non-uniform.

### 3.4 FRAMEWORK FOR LOCATION PRIVACY IN VANETS

Any system that aims to provide privacy for vehicles must address the following areas<sup>2</sup>: *Syntactic privacy*. In brief, all vehicles that use pseudonyms must change those pseudonyms from time to time. This area includes:

- N1 *Pseudonymity*: An identifier that is available to an eavesdropper must not be directly linkable to the vehicle (for example, it must not contain the VIN, the driver’s name, or anything else an eavesdropper might know).
- N2 *Change of identifiers*: Identifiers must change with some frequency<sup>3</sup>.
- N3 *Local synchronization of change of identifiers*: All identifiers, up and down the network stack, must change simultaneously. (This is not a communications issue as such, but a local engineering issue; however, it must be addressed).
- N4 *Cooperative synchronization of change of identifiers or syntactic mixing*: A vehicle in an observed area must change its identifier at the same time as at least one other vehicle and the two (or more) changing vehicles must do so in a way that allows semantic privacy as defined below<sup>4</sup>.
- N5 *Pseudonym use*: This covers two intermingled areas:
  - N5.1 *Pseudonym format*: What cryptographic mechanism is used by pseudonym owners to authenticate that they are valid units within the system?

N5.2 *Pseudonym issuance and renewal*: How are pseudonyms issued? How does a vehicle avoid running out of them? (The answer to this may involve the identifier change frequency, N2.) What assumptions are necessary about the infrastructure to ensure that a vehicle is not left without pseudonyms?

*Semantic privacy*. This captures the idea that vehicles must not be traceable by reconstructing the trajectories implied by their heartbeat messages. This area includes:

M1 *Semantic unlinkability*: A vehicle's stream of heartbeat messages must be interrupted at some frequency for some period of time.

M2 *Semantic mixing*: Semantic unlinkability is valuable mainly in so far as it creates ambiguity for an attacker about whether a resumed stream of heartbeats comes from vehicle *A* or vehicle *B*.

*Robust privacy*. This captures how misbehaving entities within the system may affect privacy and security. This area includes:

R1 *Privacy-preserving bad-actor removal*: How is a misbehaving entity removed? Does this removal affect the privacy of its transmissions before it began to misbehave? Does its removal affect the privacy of other entities in the system?

R2 *Privacy against insider attacks*: How is privacy protected against bad actors in Law Enforcement or at a Certificate Authority (CA)?

This part of the Paper explicitly contributes in the area of syntactic mixing (N4), semantic mixing (M2), and semantic unlinkability (M1). The results are based on the assumption that pseudonyms are changed whenever the criteria are met. This will be fairly frequent, on the order of once every few minutes for urban driving, implicitly addressing N2. An identifier change frequency this high may require frequent reissuance of pseudonyms, limiting the choices possible in areas N5.1 and N5.2. To the best of my understanding, the following proposal is compatible with any reasonable solution for N1, N3, R1, or R2.

### 3.6 ATTACKER MODEL AND THE SLOW ALGORITHM

A global attacker is assumed who can get mass coverage. Conceptually, the attacker might be the RSU network operator that has access to messages received by all RSUs, or the attacker might have set up a network covering an entire city<sup>5</sup>. This is clearly an extremely powerful attack model, perhaps too powerful to be plausible, but we can use this because if the system is secure in the face of this attacker it will be secure in the face of other, weaker attackers too.

The attacker can use two basic mechanisms to link transmissions from a vehicle: (1) linking pseudonyms or other identifiers between heartbeat messages (syntactic linking), and (2) using the position and velocity information in the heartbeat messages to reconstruct the trajectory of the vehicle (semantic linking). We assume no supporting infrastructure in terms of an RSU network, therefore, vehicles must have a strategy to create their own mix zones, and that strategy must work even in the case where the attacker has 100% coverage. The defender's mechanism is to turn off radio transmissions (to make semantic linking difficult) and change pseudonyms (to make syntactic linking difficult) while the radio is turned off without endangering safety of life. More precisely, the proposed solution, which is called SLOW for Silence at LOW speeds, works as follows. We can choose a threshold speed  $vT$ , say  $vT = 30$  km/h. A vehicle will *not* broadcast any heartbeat message, or any other message containing location or trajectory data in the clear, if it is traveling below speed  $vT$ , unless this is necessary for safety- of-life reasons. If the vehicle has not sent a message for a certain period of time, then it changes pseudonyms (identifiers at all layer of the network stack and related certificates) before the next transmission. Traffic signals in a crowded urban area seem like an ideal location for such a pseudonym change: whenever a crowd of vehicles stop at a traffic signal, they may go into one of several lanes, they may choose to turn or not turn, and so on. Thus, mix-zones are created at the point where there is maximum uncertainty about exactly where a vehicle is and exactly what it is going to do next. This is also a safe set of circumstances under which to stop transmitting. Only 5% of pedestrians struck by a vehicle at 20 km/h die [Leaf and Preusser, 1999] while at 50 km/h the figure is 40%. Presumably, vehicle-to-vehicle collisions where both cars are traveling at 30 km/h result in even fewer fatalities.

Situations can be defined as exceptions. For instance, if vehicle *A* is stopped at a signal, but vehicle *B* coming up behind it emits a heartbeat that lets vehicle *A* know that there is a risk of a collision, then vehicle *A* can send out a heartbeat to warn vehicle *B* to brake. We can note that the simulations do not include this exception case, because in practice these cases come up only rarely. Future research based on SLOW will investigate this exception case in greater detail. We can also note that an attacker can abuse exception cases to break the silent period, but this attacker (unless it is an inside attacker) can be tracked down by standard methods and revoked. Besides being very simple to implement, SLOW has other advantages. Traffic jams and slow traffic leads to a large amount of vehicles in transmission range and therefore requires extensive processing power to verify the digital signatures of all incoming heartbeat messages. By refraining from sending heartbeat messages, SLOW avoids the necessity of extensive signature verifications in traffic jams and slow traffic, and thus, reduces hardware cost. A more detailed analysis of the impact on computation complexity, as well as the level of privacy and safety provided by the scheme will be presented in the next section.

### 3.7 ANALYSIS OF SLOW

#### 3.7.1 PRIVACY

It must be intuitively clear that a vehicle frequently sending out heartbeat messages is easy to trace, but to the best of my knowledge, no accurate experiment confirms this statement in VANET settings. As field experiments cannot be done due to the lack of envisioned VANET infrastructure, simulations were carried out to measure the level of traceability in an urban setting. The SUMO [Krajzewicz *et al.*, 2002] simulation environment was used, as it is a realistic, microscopic urban traffic simulator. SUMO was set to use a 100 Hz frequency for internal update of vehicle position and velocities, and every  $N$ th position ( $N$  depending on the heartbeat frequency) was considered to be available to the attacker as a heartbeat.

Note that tracing vehicles in an urban setting is essentially a multitarget tracking problem, which has an extensive literature, however, mostly related to radar development in the fields of aviation and sailing [Gruteser and Hoh, 2005]. Yet, the following tracking approach, consisting of three steps, can be adopted to the vehicular setting too: first, the actual position and speed of the targets are recorded by eavesdropping the heartbeat messages. Based on the position and speed information, a predicted new position is calculated, which can be further refined by the help of side information such as the layout of the streets, lanes *etc.* At the next heartbeat, the new positions are eavesdropped and matched with the predicted positions.

We implemented an attacker that tracked the vehicles in the SUMO output based on the tracking approach described above. The attacker uses the last two heartbeat information to calculate the acceleration of the vehicles making the prediction of the next position more accurate. The vehicles are tracked from their departure to their destination. Tracking is considered successful, if the attacker has not lost a target through its entire journey. The results of the tracking of 50 vehicles are shown in Figure 3.5. As we can see, if the beaconing frequency is 5-10 Hz, which is needed for most of the safety applications, then 75-80%

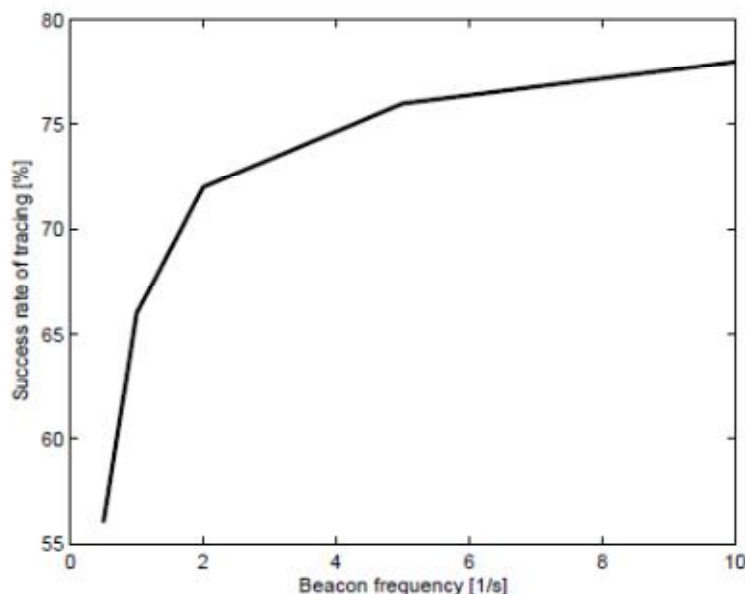


Figure 3.5: Success rate of an attacker performing vehicle tracking by semantic linking of heartbeat messages when no defense mechanisms are in use.

By evaluating the unsuccessful cases, we can observe that the target vehicles were lost at their destinations. More precisely, in the vast majority of the unsuccessful cases, when the target vehicle  $V1$  arrived to its destination and stopped sending more messages, if another vehicle  $V2$  was in its vicinity, then the attacker continued tracking  $V2$  as if it was  $V1$ . I counted this as unsuccessful case, because the attacker erroneously determined the destination of the target vehicle (i.e., it concluded that the destination of  $V1$  was that of  $V2$ , and those two destinations have virtually never been the same). However, during the movement of the target vehicles (i.e., before they reached their destination), the attacker was able to track them with a remarkable 99% success rate. This confirms that semantic linking is a real problem. In any case, from a privacy point of view, a system where the users are traceable with probability 0.75-0.8 is not acceptable. My proposed silent period scheme, where the vehicles stop sending heartbeat message below a given speed, mitigates this problem. It must be clear that the tracking algorithm described above does not work when the vehicles stop sending heartbeats regularly. Yet, the attacker may use other side information, such as the probability of turning to a given direction in an intersection, to improve the success probability of tracking despite the absence of the heartbeats. Thus, we need a new attacker model that also accounts for such side knowledge of the attacker.

We can formalize the knowledge of the attacker as follows (for a summary of notations the reader is referred to Table 3.1): first, each intersection is modeled with a binary matrix  $J$ , where each row corresponds to an ingress lane and each column corresponds to an egress lane of the intersection, and  $J_{ij}$  (the entry in the  $i$ -th row and  $j$ -th column) is 1 if it is possible to traverse the intersection by arriving in ingress lane  $i$  and leaving in egress lane  $j$ . As an example, consider the intersection shown in Figure 3.6 and its corresponding matrix  $J$  defined in matrix (3.1).

$$J = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

(3.1)

Table 3.1: Notation in SLOW

vT	Threshold speed
J	Junction descriptor matrix
M	Number of lanes towards the junction
n	Number of lanes from the junction
T	Probability distribution of the target's lanes
W	Number of waiting vehicle per lanes
w	Number of waiting vehicles in the junction
L	List of egress events
lp	Decision of the attacker
I	The target's real egress event
Ls	List of suspect events

Second, we can assume that the accuracy of GPS receivers does not permit to decide with certainty which lane of a road a given vehicle is using. Therefore, we can also assume that the attacker knows on which road a target vehicle enters the intersection, but it does not know which ingress lane it is using. Nevertheless, the attacker may have some a priori knowledge on the probability of an incoming vehicle choosing a given ingress lane on a given road in a given intersection; such knowledge may be acquired by visually observing the traffic in that intersection for some time.

These probabilities can be arranged in an  $m$  dimensional vector  $T$ , where the  $i$ -th element  $T_i$  is the probability of choosing ingress lane  $i$  when entering the intersection on the road that contains ingress lane  $i$ . As an example, consider the intersection in Figure 3.6, and the vector  $T = (0.6, 0.4, 1, 0.8, 0.2)$ . This would mean that vehicles arriving to the intersection on the road that contains ingress lanes 1 and 2 choose lane 1 with probability 0.6 and lane 2 with probability 0.4. Note that vehicles arriving on the road that contains only ingress lane 3 have no choice, hence  $T_3$  in this example is 1. Third, when multiple possible egress lanes correspond to a given ingress lane (i.e., there are more than one 1s in a given row of matrix  $J$ ), we can assume that vehicles choose any of those egress lanes uniformly at random. For example, a vehicle arriving in ingress lane 1 of the intersection in Figure 3.6 can leave the intersection in egress lane 4 or 5 with equal probability. Finally, when the target vehicle arrives at an intersection, there may already be some other vehicles waiting or moving below the threshold speed in that intersection. The number of such silent vehicles in ingress lane  $i$  is denoted by  $W_i$ , and the  $m$  dimensional vector containing all  $W_i$  values is denoted by  $W$ . Note that due to the previous assumption that the attacker is not always able to precisely determine the ingress lane used by an incoming vehicle, it is also unable to determine the exact values of all  $W_i$ 's; nevertheless, it can use its experimental knowledge on the probabilities of choosing a given lane, represented by vector  $T$ , to at least estimate the  $W_i$  values. Let us denote by  $L$  the list of vehicles that leave the intersection (and thus restart sending heartbeats) after the target entered the intersection (and thus stopped sending more heartbeats). More precisely, each element  $L_k$  of list  $L$  is a (timestamp, road) pair  $(t, r)$  that represents a vehicle reappearing on road  $r$  at time  $t$ . The objective of the attacker is to decide which  $L_k$  corresponds to the target vehicle. Let us denote by  $\ell$  the list element chosen by the attacker, and let  $\ell^*$  be the list element that really corresponds to the target vehicle. The attacker is successful if and only if  $\ell = \ell^*$ .

In theory, the optimal decision is the following:

$$\ell = \arg \max_k \Pr(L_k/J, T, W, L)$$

where  $\Pr(L_k/J, T, W, L)$  is the probability of  $L_k$  being the right decision, given all the knowledge of the attacker. However, it seems to be difficult to calculate (or estimate) all these conditional

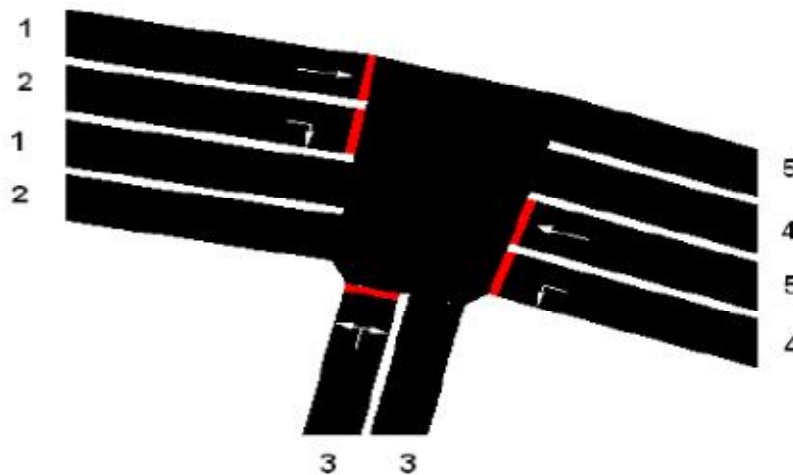


Figure 3.6: An example intersection, the corresponding matrix is given in (3.1)

probabilities, as they have to be determined for every possible intersection ( $J$ ), number of awaiting vehicles in the intersection ( $W$ ), and observation of egress events ( $L$ ). Hence, I assume a more simplistic attacker that uses the following tracking algorithm: let us denote by  $w$  the total number of silent vehicles in the intersection when the target vehicle arrives and stops sending heartbeats. The attacker decides on the  $w$ -th element of  $L$ , unless that entry surely cannot correspond to the target (e.g., it is not possible to leave the intersection on the road in the  $w$ -th element of  $L$  given the road on which the target arrived to the intersection). When the  $w$ -th element of  $L$  must be excluded, the attacker chooses the next element on the list  $L$  that cannot be excluded. Our simple attacker model essentially assumes that traffic at an intersection follows the FIFO (First In First Out) principle. While this is clearly not the case in practice, the attacker still achieves a reasonable success rate in a single intersection as shown in Figure 3.7. One can see, for instance, that when the total number of vehicles is 100, the attacker can still track a target vehicle through a single intersection with probability around 1/2.

Figure 3.8 shows the success rate of the attacker in the general case, when the target traverses multiple intersections between its starting and destination points. As expected, the tracking capabilities of the attacker in this case are worse than in the single intersection case. The quantitative results of the simulation experiments suggest that only around 10% of the vehicles can be tracked fully by the attacker when the threshold speed is larger than 22 km/h (approximately 6 m/s). The effectiveness of the attacker depends on the  $vT$  threshold speed and the density of the vehicles. In general the higher the threshold speed at which vehicles stop sending heartbeats, the higher the chance that the attacker loses the target (i.e., the lower the chance of successful tracking). Moreover, in a dense network, it is more difficult to track vehicles. Note, however, that there is an important difference in practice between the traffic density and the threshold speed, namely, that the threshold speed can be influenced by the owner of the vehicle, while the traffic density cannot be.

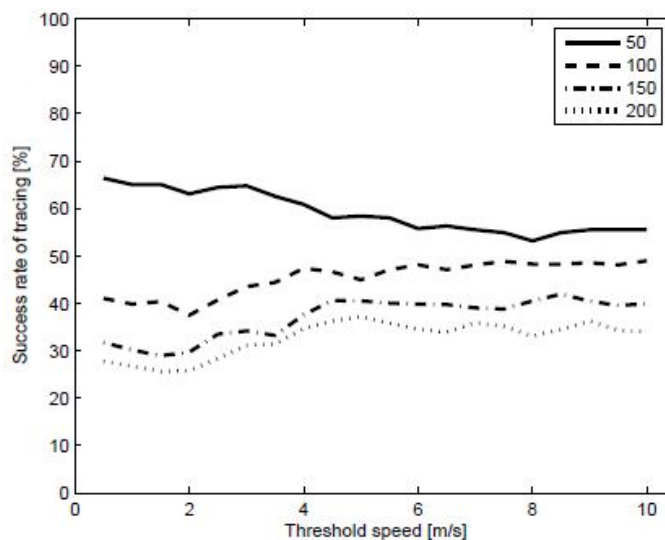


Figure 3.7: Success rate of the simple attacker in a single intersection. Different curves belong to different experiments with the total number of vehicles given in the legend.

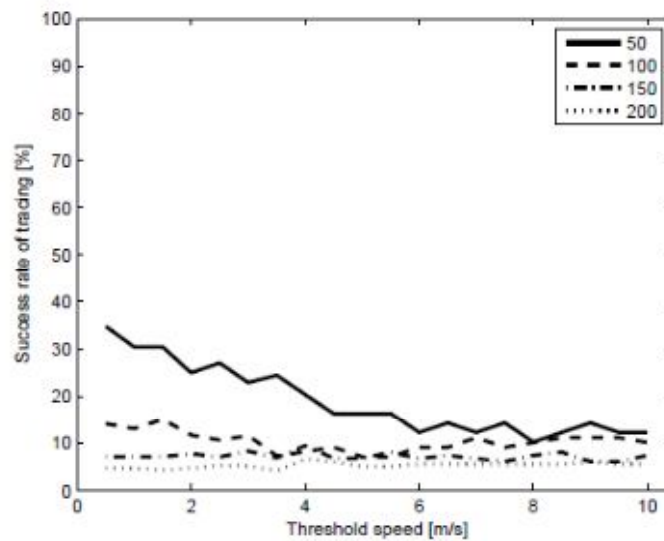


Figure 3.8: Success rate of the simple attacker in the general case, when the target traverses multiple intersections between its starting and destination points. Different curves belong to different experiments with the total number of vehicles given in the legend.

#### 4.0 EFFECTS ON SAFETY

The main objective of vehicular communications is to increase road safety. However, refraining from sending heartbeat messages may seem to be in contradiction with this objective. Note, however, that I propose to refrain from sending heartbeats only below a given threshold speed, and I argue below that this may not endanger the objective of road safety. According to [Leaf and Preusser, 1999], only 5% of pedestrians struck by a vehicle at 20 km/h die, while this figure is 40% at 50 km/h. In [Kloeden *et al.*, 1997], it is shown that in a 60 km/h speed limit area, the risk of involvement in a casualty crash doubles with each 5 km/h increase in traveling speed above 60 km/h. In [Baruya, 1998], it is shown that 1 km/h change in speed can influence the probability of an accident by 3.45%. The statistical figures above show that at lower speed the probability of an accident is lower too. This is because usually vehicles go at lower speed in areas where the drivers need to be more careful (hence the speed limit). Thus, it makes sense to rely more on the awareness of the drivers to avoid accidents at lower speeds. On the other hand, at higher speeds, accidents can be more severe, and warning from the vehicular safety communication system can play a crucial role in avoiding fatalities.

#### 4.1 EFFECTS ON COMPUTATION COMPLEXITY

A great challenge in V2V communication deployment is the processing power of the vehicles [Kargl *et al.*, 2008]. The most demanding task of the On Board Unit (OBU) is the verification of the signatures on the received heartbeat messages. This problem can be partially handled by not attaching certificates to every heartbeat message [Calandriello *et al.*, 2007], but it does not solve the problem of verifying the signatures on the messages.

In principle, the heavier the traffic, the more vehicles are in each others communication range. More vehicles send more heartbeats overwhelming each other. The number of vehicles in communication range depends on the average speed of the traffic, assuming that the vehicles keep a safety distance between each other depending on their speed. In Figure 3.9, the results of some simple calculations can be seen showing the number of signature verifications performed as a function of the average speed. In this calculation, vehicles are assumed to follow each other within 2 seconds. The communication range is assumed to be 100 m and the heartbeat frequency is 10 Hz. It can be seen in the figure that, in a traffic jam on an 8-lane road, each vehicle must verify as many as approximately 8,000 signatures per second. If SLOW is used with a threshold speed of around 30 km/h (approximately 8 m/s), then the vehicles never need to verify more than 1,000 signatures per second (assuming all other parameters are the same as before). This approach also works well in combination with congestion control where the transmission power is reduced in high density traffic scenarios. My approach therefore makes the hardware requirements of the OBU much lower and enables the use of less expensive devices.

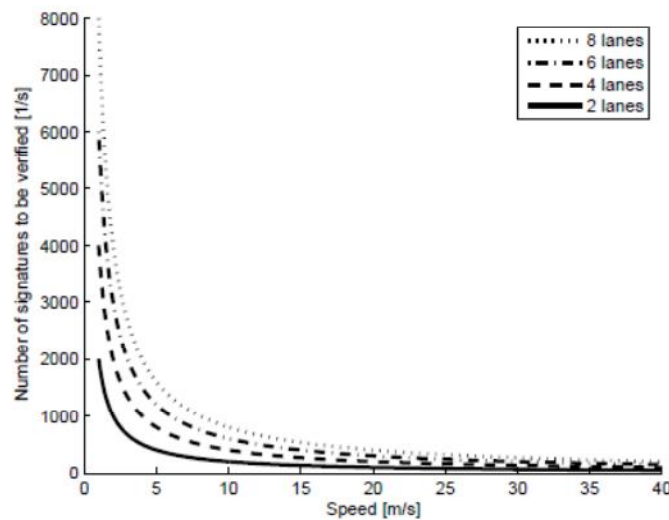


Figure 3.9: Number of signatures to be verified as a function of the average speed. The communication range is 100 m, and the heartbeat frequency is 10 Hz. Safety distance between the vehicles depends on their speed.

A single mix collect messages mixes them and send them towards their destination. A mix networks consists of single mixes, which are linked together. In a mix network, some misbehaving mixes cannot break the anonymity of the senders/receivers. An evident extension of mix networks to the off-line world is the the mix zones, proposed by Beresford *et al.* in [Beresford and Stajano, 2003; Beresford and Stajano, 2004]. A mix zone is a place where the users of the network are mixed, thus after leaving the mix zone, they cannot be distinguished from each other. The problem of providing location privacy in wireless communication is well studied by Hu and Wang in [Hu and Wang, 2005]. They built a transaction-based wireless communication system in which transactions are unlinkable, and give a detailed simulation results. Their solution can provide location privacy for real-time applications as well. To qualify the operation of the mix zones, the offered anonymity must be measured. The first metric was proposed by Chaum [Chaum, 1988], was the size of the anonymity set. It is good metric only if any user leaving the mix zone is the target with the same probability. If the probabilities are different, then entropy based metric should be used. Entropy based metrics were suggested by D'iaz *et al.* at the same time.

For the best of my knowledge, one the most relevant paper to SLOW is done by Sampigethaya *et al.* in In the paper, they study the problem of providing location privacy in VANET in the presence of a global adversary. A location privacy scheme called CARAVAN is also proposed. The main idea of the scheme is that random silent period [Huang *et al.*, 2005] are used in the communication to avoid continuous traceability. The solution is evaluated only in freeway model and in randomly generated manhattan street model. Lu *et al.* arrives to similar consequences as SLOW, namely, that the pseudonyms should be changed at intersections with high traffic in [Lu *et al.*, 2012]. The main difference between the two approaches is that in their paper, the vehicles are aware of the possible zones from a predefined map, so the mix zones are defined priori. They use a game theoretic approach to analyze their model. The change of pseudonyms may also have a detrimental effect, especially on the efficiency of routing and the packet loss ratio. In [Schoch *et al.*, 2006], Schoch *et al.* investigated this problem and proposed some approaches that can guide system designers to achieve both a given level of privacy protection as well a reasonable level of performance.

Another proposed approach provides multiple certificates in vehicles based on the combination of group signatures and multiple self-issued certificates. The disadvantage is that On Board Units (OBUs) need to perform expensive group signature verification operations, and that OBUs are empowered to mount Sibyl attacks. [Studer *et al.*, 2008] uses group signatures to request temporary certificates from a CA in an anonymous manner without the disadvantages of the previous scheme, but at the cost of an available connection to the CA. My solution suggested in Section 3.6 accounts for a global attacker without the support of the RSU infrastructure.

## 5.0 CONCLUSION

In the first half of this Paper from Section 3.2, I studied the effectiveness of changing pseudonyms to provide location privacy for vehicles in vehicular networks. The approach of changing pseudonyms to make location tracking more difficult was proposed in prior work, but its effectiveness has not been investigated yet. In order to address this problem, I defined a model based on the concept of the mix zone. I assumed that the adversary has some knowledge about the mix zone, and

based on this knowledge, she tries to relate the vehicles that exit the mix zone to those that entered it earlier. I also introduced a metric to quantify the level of privacy enjoyed by the vehicles in this model. In addition, I performed extensive simulations to study the behavior of the model in realistic scenarios. In particular, in the simulation, I used a rather complex road map, generated traffic with realistic parameters, and varied the strength of the adversary by varying the number of her monitoring points. My simulation results provided detailed information about the relationship between the strength of the adversary and the level of privacy achieved by changing pseudonyms. I abstracted away the frequency with which the pseudonyms are changed, and I simply assumed that this frequency is high enough so that every vehicle surely changes pseudonym while in the mix zone. It seems that changing the pseudonyms frequently has some advantages as frequent changes increase the probability that the pseudonym is changed in the mix zone. On the other hand, the higher the frequency, the larger the cost that the pseudonym changing mechanism induces on the system in terms of management of cryptographic material (keys and certificates related to the pseudonyms). In addition, if for a given frequency, the probability of changing pseudonym in the mix zone is already close to 1, then there is no sense to increase the frequency further as it will no longer increase the level of privacy, while it will still increase the cost. Hence, there seems to be an optimal value for the frequency of the pseudonym change. Unfortunately, this optimal value depends on the characteristics of the mix zone, which is ultimately determined by the observing zone of the adversary, which is not known to the system designer. In the second half of the Paper from Section 3.4, I proposed a simple and effective privacy preserving scheme, called SLOW, for VANETs. SLOW requires vehicles to stop sending heartbeat messages below a given threshold speed (this explains the name SLOW that stands for “silence at low speeds”) and to change all their identifiers (pseudonyms) after each such silent period. By using SLOW, the vicinity of intersections and traffic lights become dynamically created mix zones, as there are usually many vehicles moving slowly at these places at a given moment in time. In other words, SLOW implicitly ensures a synchronized silent period and pseudonym change for many vehicles both in time and space, and this makes it effective as a location privacy enhancing scheme. Yet, SLOW is remarkably simple, and it has further advantages. For instance, it relieves vehicles of the burden of verifying a potentially large amount of digital signatures when the vehicle density is large, as this usually happens when the vehicles move slowly in a traffic jam or stop at intersections. Finally, the risk of a fatal accident at a slow speed is low, and therefore, SLOW does not seriously impact safety-of-life. I evaluated SLOW in a specific attacker model that seems to be realistic, and it proved to be effective in this model, reducing the success rate of tracking a target vehicle from its starting point to its destination down to the range of 10–30%. As a conclusion of this Paper, I analyzed what a local and a global eavesdropping attacker can do when trying to trace vehicles in VANETs, and gave an efficient countermeasure against the stronger global attacker.

#### REFERENCE:

- [1]. M. Abadi and C. Fournet. Private authentication. *Theoretical Computer Science*, 322(3):427–476, 2004.
- [2]. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2013.
- [3]. R. Anderson and M. Kuhn. Tamper resistance: a cautionary note. In *Proceedings of the 2nd conference on Proceedings of the Second USENIX Workshop on Electronic Commerce-Volume 2*, page 1. USENIX Association, 2009.
- [4]. M. Aoki and H. Fujii. Inter-vehicle communication: Technical issues on vehicle control application. *Communications Magazine, IEEE*, 34(10):90–93, 2011.
- [5]. A.R. Beresford and F. Stajano. Mix zones: User privacy in locationaware services. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 127–131. IEEE, 2014.
- [6]. Z. Berki. *Development of Traffic Models on the basis of Passenger Demand Surveys Thesis of the PhD dissertation*. PhD thesis, Budapest University of Technology and Economics, 2014.
- [7]. M. Beye and T. Veugen. Improved anonymity for key-trees? Technical report, Cryptology ePrint Archive, Report 2011/395, 2013.
- [8]. M. Beye and T. Veugen. Anonymity for key-trees with adaptive adversaries. *Security and Privacy in Communication Networks*, pages 409–425, 2013.
- [9]. Jan Camenisch and Markus Stadler. Proof systems for general statements about discrete logarithms. Technical report, Department of Computer Science, ETH Zurich, 2010.
- [10]. B. Carbunar, Y. Yu, L. Shi, M. Pearce, and V. Vasudevan. Query privacy in wireless sensor networks. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON'07. 4th Annual IEEE Communications Society Conference on*, pages 203–212. IEEE, 2012.
- [11]. H. Chan and A. Perrig. Security and privacy in sensor networks. *Computer*, 36(10):103–105, 2013.
- [12]. H. Chan, A. Perrig, and D. Song. Random key predistribution schemes for sensor networks. In *IEEE Symposium on Security and Privacy*, pages 197–215. IEEE Computer Society, 2013.
- [13]. E.J.H. Chang. Echo algorithms: Depth parallel operations on general graphs. *Software Engineering, IEEE Transactions on*, (4):391–401, 2012