On the Effectiveness of Changing Pseudonyms to Provide Location Privacy in VANETs

Levente Buttyán, Tamás Holczer, and István Vajda

Laboratory of Cryptography and System Security (CrySyS)
Budapest University of Technology and Economics
{buttyan,holczer,vajda}@crysys.hu

Abstract. The promise of vehicular communications is to make road traffic safer and more efficient. However, besides the expected benefits, vehicular communications also introduce some privacy risk by making it easier to track the physical location of vehicles. One approach to solve this problem is that the vehicles use pseudonyms that they change with some frequency. In this paper, we study the effectiveness of this approach. We define a model based on the concept of the *mix zone*, characterize the tracking strategy of the adversary in this model, and introduce a metric to quantify the level of privacy enjoyed by the vehicles. We also report on the results of an extensive simulation where we used our model to determine the level of privacy achieved in realistic scenarios. In particular, in our simulation, we 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. Our simulation results provide detailed information about the relationship between the strength of the adversary and the level of privacy achieved by changing pseudonyms.

Keywords: location privacy, pseudonym, vehicular ad hoc network.

1 Introduction

Recently, initiatives to create safer and more efficient driving conditions have begun to draw strong support in Europe [4], in the US [25], and in Japan [1]. 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.

One approach to solve this problem is that the vehicles broadcast their messages under pseudonyms that they change with some frequency [18]. The change

F. Stajano et al. (Eds.): ESAS 2007, LNCS 4572, pp. 129–141, 2007.

of a pseudonym means that the vehicle changes all of its physical and logical adresses 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, changing pseudonyms is largely ineffective against a global eavesdropper that can hear all communications in the network. 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 $\boldsymbol{p}$ and moves with speed $\boldsymbol{v}$, then after some short time τ, this vehicle will most probably be at position $\boldsymbol{p}+\tau\cdot\boldsymbol{v}$. Therefore, the adversary will know that the vehicle that reports itself at (or near to) position $\boldsymbol{p}+\tau\cdot\boldsymbol{v}$ at time $t+\tau$ is the same vehicle as the one that reported itself at position $\boldsymbol{p}$ at time t, even if in the meantime, the vehicle changed pseudonym.

On the other hand, the assumption that the adversary can eavesdrop all communications in the network is a very strong one. In practice, 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. Our goal in this paper is to characterize this chance as a function of the strength of the adversary (i.e., its monitoring capabilities). In particular, our main contributions are the following:

- We define a model in which the effectiveness of changing pseudonyms can be studied. We emphasize that while changing pseudonyms has already been proposed in the literature as a countermeasure to track vehicles [18], to the best of our knowledge, the effectiveness of this method has never been investigated rigorously in this context. Our model is based on the concept of the *mix zone*. This concept was first introduced in [2], but again, to the best of our knowledge, it has not been used in the context of vehicular networks so far. We characterize the tracking strategy of the adversary in the mix zone model, and we introduce a metric to quantify the level of privacy provided by the mix zone.
- We report on the results of an extensive simulation where we used our model to determine the level of privacy achieved in realistic scenarios. In particular, in our simulation, we 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, our simulation results confirm that the level of privacy decreases as the strength of the adversary increases. However, in addition to this, our simulation results provide detailed information about the relationship between the strength of the adversary and the level of privacy achieved by changing pseudonyms.

The organization of the paper is the following: In Section 2, we introduce the mix zone model, we define the behavior of the adversary in this model, and we introduce our privacy metric. In Section 3, we describe our simulation setting and the simulation results. Finally, we report on some related work in Section 4, and conclude the paper in Section 5.

2 Model

2.1 The Concept of the Mix Zone

We consider a continuous part of a road network, such as a whole city or a district of a city. We 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, we divide the road network 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 in Part (a) of Figure 1.

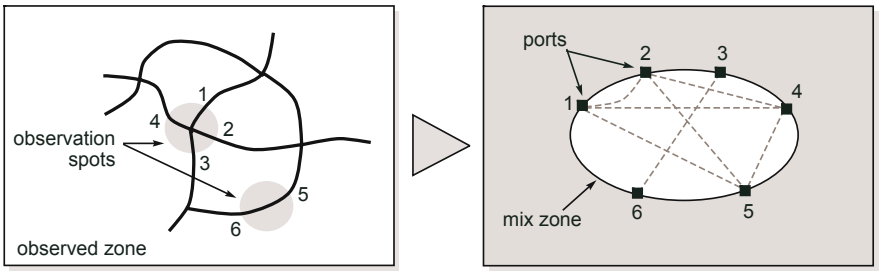

Fig. 1. Part (a) illustrates how a road network is divided into an observed and an unobserved zone in our model. In the figure, the observed zone is grey, and the unobserved zone is white. The unobserved zone functions as a *mix zone*, because the vehicles change pseudonyms and mix within this zone making it difficult for the adversary to track them. Part (b) 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 assume that the vehicles continuously change their pseudonyms[1]. In this paper,

[1] Otherwise, if the vehicles knew when they are in the unobserved zone, then it would be sufficient to change their pseudonyms only once while they are in the unobserved zone.

we abstract away the frequency of the pseudonym changes, and we simply assume that it is high enough so that every vehicle surely changes pseudonym while in the unobserved zone. We intend to relax this assumption in our future work.

Since the vehicles change pseudonyms while in the unobserved zone, that zone functions as a *mix zone* for vehicles (see Part (b) of Figure 1 for illustration). A mix zone [2,3] is similar to a mix node of a mix network [6], 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 our 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 we made in the Introduction, we assume that the adversary can track the physical location of the vehicles while they are in the 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 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, in Part (b) of Figure 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 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 assume that time is slotted, that is why $f_{ij}(t)$ is a discrete function. We 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 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 our model, the adversary picks a vehicle v in the observed zone and tracks its movement until it enters the mix zone. In the following, we 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., $\Pr\{t_{out} < T\} = 1 - \epsilon$, where ϵ is a small number, typically, in the range of $0.005 - 0.01$, and t_{out} 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 Appendix 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 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 our 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 note that the level of privacy is often measured using the anonymity set size as the metric [5], however, in our case, this approach cannot be used. The problem is that as described above, with probability ϵ, 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 a 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, this is not the case in our model.

Obviously, the success probability of the adversary is very difficult to determine analytically due to the complexity of our model. Therefor, we 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 Simulations

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

3.1 Simulation Settings

The simulation was carried out in three main phases. In the first phase, we generated a realistic map, where the vehicles moved during the simulation. This map was generated by MOVE [15], a tool that allows the user to quickly generate realistic mobility models for vehicular network simulations. Our map is illustrated in Figure 2. In fact, it is a simplified map of Budapest, the capital of Hungary, and it contains the main roads of the city . We 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 [24], which is an open source micro-traffic 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, we 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, we made the source code available on-line at http://www.crysys.hu/∼holczer/ESAS07.

We implemented the adversary as follows. First, we defined the observation spots (position and radius) of the adversary in a configuration file. Then, we 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. Once the adversary's knowledge is built, she could use that for making decisions as described above in Section 2. We executed several simulation runs in order to get an estimation for the success probability of the adversary.

We made experiments with adversaries of different strength, where the strength of the adversary depends on the number of her eavesdropping receivers.

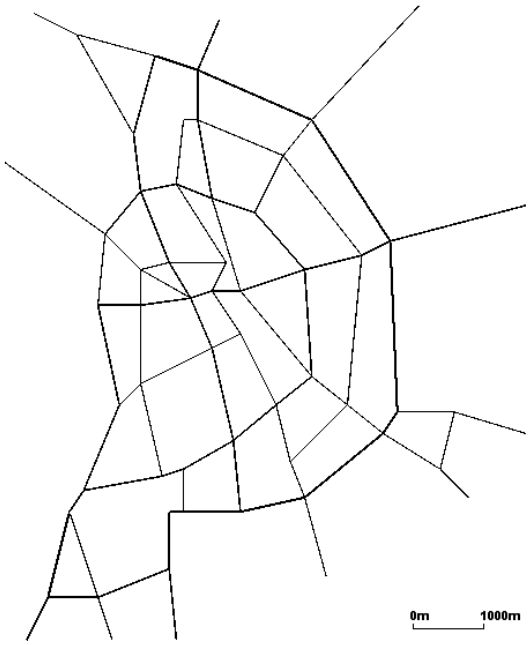

Fig. 2. Simplified map of Budapest generated for the simulation

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, we chose the first ten junctions with the largest traffic).

In addition to the strength of the adversary, we varied the intensity of the traffic. More specifically, we 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) we ran 100 simulations.

3.2 Simulation Results

Figure 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

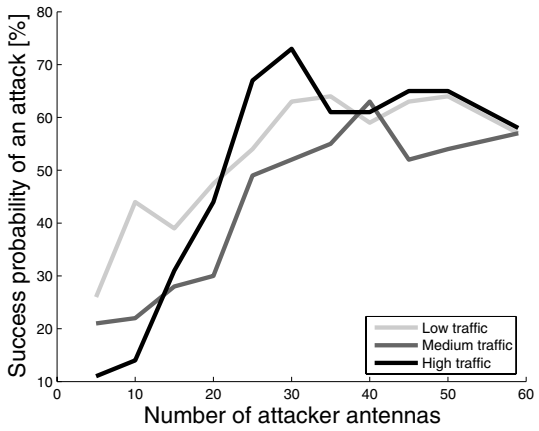

Fig. 3. Success probabilities of the adversary as a function of her strength. The three curves represent three different scenarios (the darker the line, the more intensive the traffic).

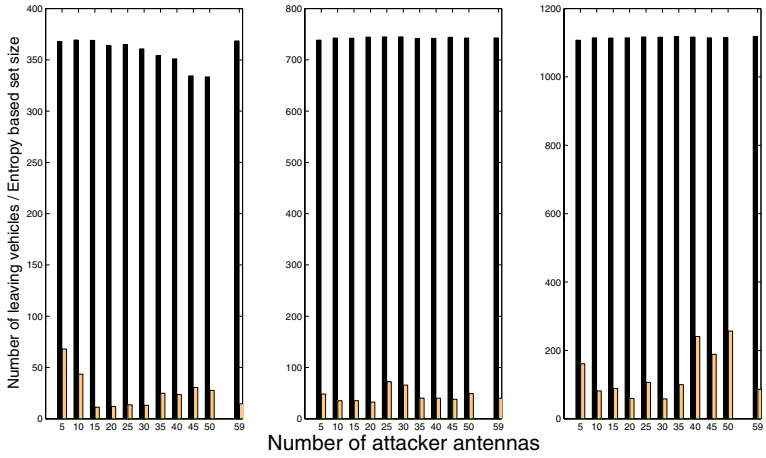

Fig. 4. The dark bars show how the size of the set V of the vehicles that exit the mix zone during the observation period varies with the strength of the adversary. The three sub-figures are related to the three different traffic situations (low traffic – left, medium traffic – middle, high traffic – 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.

can not 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..

The dark bars in Figure 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, we 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 distribution corresponding to the members of V is highly non-uniform. This is the reason why the adversary can be successful.

4 Related Work

The privacy of VANET's is a recent topic. Many author addressed the whole problem in some papers (for example in [9,14,18,19]). The problem of providing location privacy for VANET's is categorised in [10], into classes. The difference between the classes is the goal and the strength of the attacker. In [7], Choy, Jakobsson and Wetzel investigates how to obtain a balance between privacy and audit requirements in vehicular networks using only symmetric primitives.

Many privacy preserving techniques are suggested for on-line transactions (for example in [5,11]). Mainly they are based on mix networks [16,20], which was basically proposed by Chaum in 1981 [6]. A single mix collect messages mixes them and send them towards their destination. A mix networks consits of single mixes, which are linked together. In a mix network, some misbehaving mixes can not break the anonimity 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 [2,3]. A mix zone is a place where the users of the network are mixed, thus after leaving the mix zone, they can not be distinguished from each other.

The problem of providing location privacy in wireless communication is well studied by Hu and Wang in [12]. 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 anonomity must be measured. The first metric was proposed by Chaum [5], was the size of the anonimity 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 Díaz et. al [8] and Serjantov *et al.* [23] at the same time.

For the best of our knowledge, the most relevant paper to this work is done by Sampigethaya *et al.* in [21]. 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 [13] are used in the communication to avoid continous traceability. The solution is evaluated only in freeway model and in randomly generated manhattan street model.

The change of pseudonyms may also have a detrimental effect, especially on the efficiency of routing and the packet loss ratio. In [22], Schoch *et al.* investigated this problem and proposed a some approaches that can guide system designers to achieve both a given level of privacy protection as well a reasonable level of performance.

5 Conclusion and Future Work

In this paper, we 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, we defined a model based on the concept of the mix zone. We 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. We also introduced a metric to quantify the level of privacy enjoyed by the vehicles in this model. In addition, we performed extensive simulations to study the behavior of our model in realistic scenarios. In particular, in our simulation, we 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. Our simulation results provided detailed information about the relationship between the strength of the adversary and the level of privacy achieved by changing pseudonyms.

In this paper, we abstracted away the frequency with which the pseudonyms are changed, and we simply assumed that this frequency is high enough so that every vehicle surely changes pseudonym while in the mix zone. In our future work, we intend to relax this simplifying assumption, and we want to study how the level of privacy depends on the frequency of the pseudonym changes. 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 our future work, we want to characterize this dependence in more details.

Acknowledgements

This work has partially been supported by the European Commission through the SeVeCom Project (IST-027795), by the Hungarian Scientific Research Fund (T046664), and by the Mobile Innovation Center, Hungary (www.mik.bme.hu).

References

1. Advanced Safety Vehicle Program, `http://www.ahsra.or.jp/demo2000/eng/demo_e/ahs_e7/iguchi/iguchi.html`
2. Beresford, A.R., Stajano, F.: Location privacy in pervasive computing. IEEE Pervasive Computing 3(1), 46–55 (2003)
3. Beresford, A., Stajano, F.: Mix Zones: User privacy in location-aware services. In: Proceedings of First IEEE International Workshop on Pervasive Computing and Communication Security (PerSec) 2004, a workshop in PerCom (2004)
4. Communications for eSafety `http://www.comesafety.org/`
5. Chaum, D.: The Dining Cryptographers Problem: Unconditional sender and recipient untraceability. Journal of Cryptology 1(1), 65–75 (1988)
6. Chaum, D.: Untraceable electronic mail, return addresses, and digital pseudonyms, Communications of the ACM, vol. 4 (February 981)
7. Choi, J.Y., Jakobsson, M., Wetzel, S.: Balancing Auditability and Privacy in Vehicular Networks. In: Proceedings of International Workshop on QoS and Security for Wireless and Mobile Networks (Q2SWinet 2005), ACM Press, New York (2005)
8. Díaz, C., Seys, S., Claessens, J., Preneel, B.: Towards measuring anonymity. In: Dingledine, R., Syverson, P.F. (eds.) PET 2002. LNCS, vol. 2482, pp. 54–68. Springer, Heidelberg (2003)
9. Doetzer, F.: Privacy issues in vehicular ad hoc networks. In: Workshop on Privacy Enhancing Technologies, Cavtat, Croatia (May 2005)
10. Gerlach, M.: Assessing and Improving Privacy in VANETs. In: ESCAR, Embedded Security in Cars (2006)
11. Gülcü, C., Tsudik, G.: Mixing E-mail With Babel. In: Proceedings of the Network and Distributed Security Symposium - NDSS '96, February 1996, pp. 2–16. IEEE Computer Society Press, Los Alamitos (1996)
12. Hu, Y.C., Wang, H.J.: A Framework for Location Privacy in Wireless Networks. In: Proceedings of the ACM SIGCOMM Asia Workshop 2005, April 2005, ACM, Bejing, China (2005)
13. Huang, L., Matsuura, K., Yamane, H., Sezaki, K.: Enhancing Wireless Location Privacy Using Silent Period. In: IEEE Wireless Communications and Networking Conference (WCNC 2005), IEEE Computer Society Press, Los Alamitos (2005)
14. Hubaux, J.P., Čapkun, S., Luo, J.: The security and privacy of smart vehicles. IEEE Security and Privacy 4(3), 49–55 (2004)

15. Karnadi, F., Mo, Z., Lan, K.: Rapid Generation of Realistic Mobility Models for VANET. In: International Conference on Mobile Computing and Networking (ACM MOBICOMM 2005), ACM Press, New York (2005)
16. Kesdogan, D., Egner, J., Büschkes, R.: Stop-and-Go MIXes: Providing Probabilistic Anonymity in an Open System. In: Aucsmith, D. (ed.) Information Hiding. LNCS, vol. 1525, Springer, Heidelberg (1998)
17. Loca Project, http://www.loca-lab.org
18. Raya, M., Hubaux, J.P.: In: Proc. of Third ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN 2005), Alexandria (November 2005)
19. Raya, M., Hubaux, J.P.: Securing Vehicular Ad Hoc Network (Special Issue on Security of Ad Hoc and Sensor Networks). Journal of Computer Security 15(1), 39–68 (2007)
20. Reiter, M., Rubin, A.: Crowds: Anonymity for Web Transactions. ACM Transactions on Information and System Security, 1 (1998)
21. Sampigethaya, K., Huang, L., Li, M., Poovendran, R., Matsuura, K., Sezaki, K.: Caravan: Providing location privacy for VANET. In: ESCAR 2005. Proc. of 3rd workshop on Embedded Security in Cars, Cologne, Germany (2005)
22. Schoch, E., Kargl, F., Leinmüller, T., Schlott, S., Papadimitratos, P.: Impact of Pseudonym Changes on Geographic Routing in VANETs. In: Buttyán, L., Gligor, V., Westhoff, D. (eds.) ESAS 2006. LNCS, vol. 4357, Springer, Heidelberg (2006)
23. Serjantov, A., Danezis, G.: Towards an information theoretic metric for anonymity. In: Dingledine, R., Syverson, P.F. (eds.) PET 2002. LNCS, vol. 2482, Springer, Heidelberg (2003)
24. SUMO Simulation of Urban MObility, http://sumo.sourceforge.net/
25. Vehicle Safety Communications Project http://www-nrd.nhtsa.dot.gov/pdf/nrd-12/CAMP3/pages/VSCC.htm

A Appendix

In this appendix, we show that the decision algorithm of the adversary described in Subsection 2.3 realizes a Bayesian decision. We use the following notations:

- 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, we always use the one which makes the presentation simpler.
- $k \in 1 \ldots M \cdot T$, where M is the number of ports, and T is the length of the attack measured in timeslots.
- $C = [c_k]$ is a vector, where c_k 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., $N = \sum_{k=1}^{MT} c_k$).
- $p_s(k)$ 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 $p_s(k)$ can be computed as: $p_s(k) = q_{sj} f_{sj}(t)$, 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: $p(k) = \sum_{s=1}^{M} p_s(k)$.
- $\Pr(k|C)$ 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 C.

We want to 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:

$$\Pr(k|C) = \frac{\Pr(C|k)p_s(k)}{\Pr(C)}$$

Then k^* can be computed as:

$$k^* = \arg\max_k \frac{\Pr(C|k)p_s(k)}{\Pr(C)} = \arg\max_k \Pr(C|k)p_s(k)$$

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

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

$\Pr(C|k)$ can be multiplied and divided by $\frac{p_k}{c_k}$ to simplify the equation:

$$\Pr(C|k) = \frac{c_k}{p_k} \left(\frac{N!}{c_1! \dots c_{MT}!} \prod_{j=1}^{MT} p(j)^{c_j} \right)$$

where the bracketed part is a constant, which does not have any effect on the maximization, thus it can be omitted.

$$k^* = \arg\max_k \frac{c_k}{p_k} p_s(k) = \arg\max_k \frac{c_k}{p_k N} p_s(k) = \arg\max_k \frac{\widehat{p_k}}{p_k} p_s(k)$$

where $\widehat{p_k}$ is the empirical distribution of k (i.e., $\widehat{p_k} = c_k/N$). If the number of vehicles in the mix zone is large enough, then $\frac{\widehat{p_k}}{p_k} \approx 1$. Thus correctness of the intuitive algorithm described in Subsection 2.3 holds:

$$k^* = \arg\max_k p_s(k)$$

This means that if many vehicles are travelling in the mix zone, then the attacker must choose the vehicle with the highest $p_s(k)$ probability.