Connected Vehicles: Surveillance Threat and Mitigation

Jonathan Petit[†*]
jpetit@securityinnovation.com

Djurrre Broekhuis[†],
Michael Feiri[†]
djurreb@gmail.com
m.feiri@utwente.nl

Frank Kargl[†‡]
frank.kargl@uni-ulm.de

[*]Security Innovation
Wilmington, MA
United States

[†]Services, Cybersecurity and Safety
University of Twente
The Netherlands

[‡]Institute of Distributed Systems
University of Ulm
Germany

ABSTRACT

Intelligent Transportation Systems (ITSs) are an upcoming technology that allow vehicles and road-side infrastructure to communicate to increase traffic efficiency and safety. To enable cooperative awareness, vehicles continually broadcast messages containing their location. These messages can be received by anyone, jeopardizing location privacy. A misconception is that such attacks are only possible by a global attacker with extensive resources (e.g. sniffing stations at every intersections giving a full city-wide coverage). In this paper, we demonstrate the feasibility of location tracking attack in an ITS in the presence of a mid-sized attacker (i.e. an attacker that has partial network coverage but can choose which parts to cover). We conduct an empirical study on the campus of the University of Twente by deploying ITS hardware on a small scale. As road intersections are likely targets for an attacker to eavesdrop, we propose a graph-based approach to determine which intersections an attacker should cover. We also derive a cost analysis that gives an indication of the financial resources an attacker needs to track a vehicle. To mitigate location tracking attacks, we assess the benefit of pseudonym change strategies and propose a privacy metric to quantify a vehicle's level of privacy in the presence of mid-sized attackers. Experiment results demonstrate that tracking is feasible even if such an attacker covers a small number of intersections. For example, with only two sniffing stations, a mid-sized attacker can track the target vehicle on a zone-level 78% of the time, and on a road-level 40% of the time. Pseudonym schemes harden tracking by increasing the number of sniffing stations required.

Keywords

Privacy, Intelligent Transportation Systems, Tracking Attack

1. INTRODUCTION

Modern vehicles are becoming increasingly equipped with a multitude of sensors that allow them to gather data on their surroundings. Vehicles may, for example, collect information about the temperature, road conditions or the distance to other objects and vehicles. Along with these sensors, vehicles are also starting to become equipped with wireless communication systems that allow them to communicate with other vehicles and infrastructure and set up Vehicular Ad-Hoc Networks (VANETs). Combining these two features allows for cooperative awareness and the development of advanced applications. These networked, context-aware vehicular networks along with their supporting infrastructure are often called Intelligent Transportation Systems (ITSs).

ITS applications can significantly improve driver safety and comfort, for example by providing warnings on road dangers or traffic jams, or automatically braking a vehicle when a collision seems likely. At the same time, vehicles collecting and sharing data about themselves and their surroundings gives rise to privacy issues. Many envisioned ITS applications rely on vehicles knowing the position of both themselves and their neighbours. Therefore, as part of cooperative safety applications, real-time location and trajectory beacons are periodically broadcast. Without privacy protection, broadcasting these beacons jeopardises the location privacy of drivers by allowing them to be tracked. Current standards acknowledge the issue of privacy, and recommend the use of short-term credentials (named *pseudonyms*) instead of long-term credentials. However, detail regarding pseudonym management, and especially pseudonym change strategies, are still lacking.

A general misconception is that location tracking attacks are only accessible to global observers with extensive resources (e.g. governmental agencies, prevalent companies). But the deployment of ITSs will put radio networking equipment into the hands of the general public. This allows anyone to eavesdrop vehicle-to-x communications, and thus, to track vehicles. Tracking may be of particular interest to criminals when we consider certain classes of vehicles, such as police vehicles or money transports. For example, if burglars could track patrolling police vehicles they can wait until all police vehicles are outside of a certain area before attempting a robbery, which would increase the response time before the police can be at the crime scene to intervene. Besides the criminal use-case, the ability to track specific individuals in real-time is attractive for businesses, insurance companies, or any curious citizen. We introduce a new type of attacker, named the *mid-sized attacker*, who can install sniffing stations at specific locations to perform tracking attacks.

In this paper we investigate empirically the feasibility of location tracking attacks in an intelligent transportation system, by deploying ITS equipment on the campus of the University of Twente. To our knowledge this is the first real-world experiment that demonstrates location privacy attack in VANETs. Results show that deploying two sniffing stations is sufficient to give 78% zone-level tracking (i.e. identify if the security guard vehicle is in the residential or business area of the campus). By determining the requirements and resources of an attacker, a cost model is derived, giving a realistic overview of the likelihood of privacy attacks when ITS will be deployed. Finally, to mitigate tracking, we assess the benefit of different pseudonym change strategies by using a new privacy metric that captures the mid-sized attacker capabilities. We conclude that even though pseudonyms cannot eliminate the risk of tracking completely, they can still form an important line of defence. Through this paper we hope to shed light on the complexities of location privacy in vehicular networks, and more importantly, to raise awareness of the need to ensure such privacy in all upcoming ITSs.

The rest of this paper is organised as follows: Section 2 puts our research into context by examining the related work. Section 3 gives a description of the system model, describing what components constitute an ITS. Section 4 details the attacker model considered in this paper. Especially, it introduces the mid-sized attacker, which is in our opinion the most realistic type of attacker. Section 5 introduces the new privacy metric used to capture the mid-sized attacker capabilities. Section 6 describes the experiment setup, identifies the best candidate locations for sniffing stations placement using a graph-based approach, and describes the data collection. Section 7 applies the hybrid privacy loss function to the experimental data to show the benefits of pseudonym change and derive a cost model. Section 8 concludes the paper and outlines future work. The paper is also complemented with appendices, which present results of zone-level tracking (Appendix A) and road-level tracking (Appendix B).

2. RELATED WORK

Location privacy is defined as a special type of information privacy which concerns the claim of individuals to determine for themselves when, how and to what extent location information is communicated to others [3]. Three negative effects associated with a failure to protect location privacy are location-based spam, personal well-being and safety, and intrusive inferences. The latter is most relevant to the issue of tracking, as being able to identify at which times a person is at which locations allows for inferences of, for example, a person's political views, state of health, or personal preferences. To protect location privacy, Duckham and Kulik proposed technical solutions such as anonymity, pseudonymity and obfuscation [3]. In ITS, anonymity and obfuscation of location data might drastically reduce the data quality, and thus jeopardize the cooperative awareness. Therefore, pseudonymity (i.e. using short-term identifier) is the only solution considered in current standards. A comprehensive survey of pseudonym schemes in Vehicular Ad-hoc Networks (VANET) has been done in [11].

However, using a single short-term identifier still allows linking of consecutive location samples to each other, and through this, even to an individual. For example, Hoh et al. analysed one week of pseudonymised GPS traces from drivers in Detroit, and their home-finding algorithm was able to find home locations for 85% of the drivers [6], with a median accuracy of 61 metres [8]. Using a reverse white pages lookup, correct identification of driver's home address reached 13% and names 5%. To decrease this pseudonym linkability, pseudonyms should be changed during a vehicle's trip [15].

With the many different proposed pseudonym change strategies [11], it is important to consider what trade-offs come with introducing pseudonyms into an ITS. Lefèvre et al. investigated the effects of pseudonym change strategies on an intersection collision avoidance (ICA) system [9]. They analysed the effects of three pseudonym strategies using the rate of missed accident interventions, the rate of avoided collisions, and the rate of failed interventions. They found that silent periods longer than 2 seconds strongly affect ICA applications, and that the adaptive approach only authorised average of 10 percent of pseudonym changes when the silent period was larger than 2 seconds. This indicates that whilst pseudonym changes and silent periods may be beneficial for location privacy, they may also have an impact on the main functionalities of an intelligent transportation system.

To determine how effective changing pseudonyms are, Buttyan et al. defined all areas that are unobserved by an adversary as a mix-zone [2]. As vehicles do not know when they are in a mix-zone, pseudonyms are constantly changed. They assumed that this rate of change is high enough that pseudonyms are changed at least once per mix-zone. The adversary strength was varied by eavesdropping on the k busiest junctions, with an eavesdropping range of 50 meters. Different traffic densities were simulated, and the success of the adversary was quantified by calculating the number of successful tracking attempts. Tracking was considered successful when a vehicle entering a mix-zone was correctly linked to a vehicle exiting that mix-zone. Linking was done using a basic dead reckoning approach where the probability of linking the correct vehicle was based on the speed and distance covered in the mix-zone. They found that tracking was successful 60% of the time with 30 eavesdroppers. However, they did not conduct real experiment to prove the feasibility of such attack and did not investigate pseudonym change strategies.

Humbert et al. [7] considered the problem of deploying mix zones in the presence of a passive adversary equipped with a limited number of eavesdropping stations. They proposed a game-theoretic model to evaluate the strategic behaviors of players in such tracking games. In the incomplete information case (which corresponds to the case considered in this paper), they noticed that mobile nodes' strategy highly depends on their belief about the type of adversary. Their results quantified how the lack of information by mobile nodes about the attacker's strategy leads to a significant decrease in the achievable location privacy level at Bayesian Nash Equilibrium. Their work enables system designers to predict the strategy of a local adversary and mobile nodes with limited capabilities in tracking games. However, this abstract work did not conclude on a cost model (number of eavesdropped intersections w.r.t. privacy level) nor compare different pseudonym change strategies.

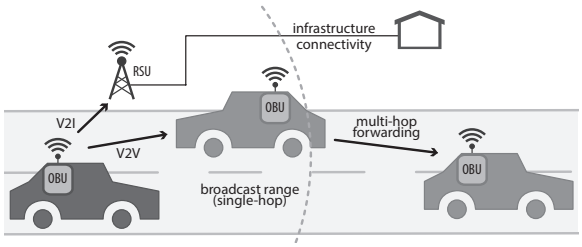

Figure 1: Typical ITS setup

Shokri et al. [13] evaluated location-based applications that expose users location to curious observers, who might collect this information for various monetary or malicious purposes. They proposed a formal framework for quantifying location privacy in the case where users expose their location sporadically. Indeed, in most location-based applications, users expose their location in a sporadic manner as opposed to a continuous manner. In the context of ITS, we can consider *'not using pseudonym'* as continuous exposition, while *'using pseudonyms'* provides a sporadic exposition (during the time frame the same pseudonym is used). In [13], location privacy is preserved by using anonymization, location obfuscation, or broadcast of fake location beacons. As described earlier, these mechanisms might have undesired effects on safety systems that require cooperative awareness [9]. Therefore, we do not consider them in this paper. Moreover, we differentiate from this work by investigating pseudonym changes and considering a mid-sized observer.

3. SYSTEM MODEL

We consider a VANET consisting of both vehicles and supporting road-side infrastructure. An example of such a set up can be seen in Figure 1. To allow vehicles in a VANET to send and receive messages, they are equipped with a station called an On-Board Unit (OBU). An OBU typically consists of a car computer with networking hardware. An OBU can collect diverse sensor information such as vehicle trajectory data or road conditions, and process and send these data. Apart from the OBUs in the vehicles, there is also static infrastructure to improve data dissemination and to provide connectivity with back end systems. These static infrastructure stations consist of Road-Side Units (RSUs), which are similar to OBUs except that they are fixed in place and typically have additional network access.

Each participating vehicle will broadcast to its immediate neighborhood (i.e. one hop) at least once per second a position beacon named Cooperative Awareness Message (CAM) [1] in Europe and Basic Safety Message (BSM) [12] in US. In this work we used CAM, but results are also valid for BSM. A typical CAM includes the unencrypted latitude and longitude of a vehicle, its trajectory, speed, a timestamp and an identifier.

Vehicles use pseudonyms to increase unlinkability between messages transmitted from the same vehicle. A pseudonym is a unique identifier with which a vehicle can communicate to other vehicles and RSUs. Pseudonyms can be changed according to a *pseudonym change strategy*. In the context of VANETs, a pseudonym change should affect all public information that can be directly linked to a vehicle. This means that all identifiers on the communication stack, such as a vehicle's MAC or IP address, as well as any (public) keys that the vehicle uses for authentication should be changed. Vehicles will, however, probably only have a limited number of pseudonyms available to them, and generating new pseudonyms too often is not feasible. This is especially true as access to pseudonyms needs to be limited so that vehicle cannot use multiple pseudonyms at the same time, as this would confuse ITS applications and open up the possibility of sybil attacks [17]. To reduce this problem, it is logical to use the same pseudonym for multiple messages before the pseudonym is changed. To stop an attacker linking old and new pseudonyms, a pseudonym change should be spatially and temporally coordinated amongst different vehicles. In general this means the pseudonym changes are only effective if there are enough neighbours to confuse an attacker, and these neighbours need to change pseudonyms around the same time, resulting in collaborative pseudonym changes.

4. ATTACKER MODEL

Within the attacker model defined in [10], we define an attacker targeting the communication channel as follows:

- *Scope*: The scope is the area over which the attacker can eavesdrop. On one end of the spectrum is the *global attacker*, which has complete coverage and can eavesdrop on any message that has been transmitted in the network. On the other end of the spectrum is the *local attacker*. This is an attacker that can only cover one small area. In between these two extremes, there is the *mid-sized attacker* (MA), which can choose to cover any number of different local areas without obtaining complete network coverage.
- *Passive/Active*: A passive attacker is only capable of receiving and processing any packets that it receives, whereas an active attack can also inject packets into the network.
- *Internal/External*: An internal attacker possesses keys and credentials that make it a legitimate participant of the system, whereas an external attacker does not.
- *Tracking Period*: The tracking period defines over what period an attacker tries to link location samples and track a vehicle. We distinguish between the following:
  - *Short-term tracking* means that an attacker tries to link consecutive location samples occurring in a time frame of a couple of seconds. Given multiple location samples of different vehicles, the attacker tries to link the location samples to the specific vehicles that sent them.
  - *Mid-term tracking* means that an attacker tries to link position samples from a single trip. A vehicle trip is the entire time period from when a vehicle start a journey until it ends, and can be in the order of a couple of minutes to a couple of hours.
  - *Long-term tracking* means that not only does an attacker try to link consecutive location samples, but it is also tries to link different sets of location samples from different trips. Long-term tracking can cover a time period of over one day. For example, the attacker tries to identify that a police vehicle that was tracked in a certain area one day

is the same vehicle that passes through that area the next day or a couple of days later.

- *Road-level/Zone-level Tracking*: The tracking level is the level of granularity that a vehicle is tracked. Road-level tracking means that an attacker knows which road a vehicle is on. Zone-level tracking means that an attacker knows which zone, or set of roads, a vehicle is on.

As we are interested in location privacy attacks, we consider in this paper a mid-sized, passive, internal/external, mid/long-term tracking at road/zone level, type of attacker. This attacker can eavesdrop on CAMs transmitted by vehicles by deploying static *sniffing stations* at road intersections. The risk of tracking is then relative to the capabilities of an attacker in linking pseudonyms by (physically) identifying the target vehicle, for instance, manually by the attacker or automatically by radio fingerprinting, fingerprinting based on the contents of CAMs, or with cameras and object recognition techniques. These extra methods of re-identification might increase the resources needed by an attacker to track a vehicle. In Section 6, we describe how an attacker can determine which intersections to cover. Any intersection which is covered by a sniffing station is called an *observed intersection*. As the MA coverage depends on her available resources (e.g. number of sniffing stations she can deploy), we also investigate how the number of sniffing stations deployed impact the trackability.

5. PRIVACY METRICS

In order to compare the privacy level of a vehicle in different scenarios, a quantitative measure of privacy is needed. Privacy metrics allow us not only to assess when and where a vehicle's privacy is under threat, but also to compare the effectiveness of different mitigation strategies that aim to increase privacy. However, we consider an attacker that cannot observe the entirety of the network, which affects the applicability of traditional privacy metrics [7]. Thus, existing metrics are adapted to reflect the mid-sized attacker.

5.1 Hybrid Privacy Loss Function

When a vehicle drives through an area with only few observed intersections and a high traffic density, it is already intuitive that it experiences a higher level of privacy than when the same vehicle travels through an area with many observed intersections. As a vehicle can transition between these two situations while it is moving, it makes sense not to look at a single value of privacy for a vehicle, but to define a function that describes how the level of privacy changes over time. Such functions are typically called location privacy loss functions, and model how a vehicle loses and gains privacy over time. For example, Freudiger et al. proposed a loss function that describes the amount of privacy lost in relation to the time, the time since the last pseudonym change and the duration of the silent period [4]. In this model, privacy loss is set to zero after a pseudonym change and during the subsequent silent period. After this, the level of privacy loss increases according to a sensitivity parameter λ, which models the tracking power of the attacker. Privacy loss can increase to a set maximum, which is dependent on the number of vehicles involved in the last pseudonym change. Whilst this privacy loss function does give a good indication of the level of privacy that a vehicle has due to pseudonym changes, it does this under the assumption of a global attacker. This means that outside of silent periods, the level of location privacy is always decreasing. A mid-sized attacker is however weaker, and thus privacy does not always decrease. In fact, privacy will only decrease when a vehicle is within an observed area. Outside of these observed areas the level of privacy will stay the same, or may even increase due to other sources of uncertainty. Thus, the privacy loss function described above is not directly applicable to our scenario.

To account for these limitations, we propose an adjusted privacy loss function that takes into account the mid-sized attacker characteristics. This hybrid privacy loss function takes into account the following three sources of uncertainty in the presence of an MA: 1) uncertainty about which vehicle is the target vehicle; 2) uncertainty about the road section a vehicle is on; 3) uncertainty about the exact location of a vehicle on a road section.

The first source of entropy can be gained by collaboratively changing pseudonyms with other vehicles on the same road section, as described in Section 3. The second source of uncertainty is the different routes that a vehicle can take between intersections. This uncertainty is not present if the attacker observes adjacent intersections, as then, there is only a single possible route. This uncertainty is however present as soon as a vehicle crosses an unobserved intersection. The final source of uncertainty is the estimation error that an attacker makes on the travel time of a target between intersections. The longer the road between intersections, the more difficult it becomes to accurately predict where the target vehicle is on that road. Thus the longer a vehicle drives without being observed, the greater this uncertainty is.

A vehicle can gain privacy through any of these three sources. However, a vehicle can also lose privacy. In our model, this occurs when a vehicle encounters an observed intersection. Thus, when a vehicle encounters many observed intersections, its overall level of privacy will be lower than when it encounters many unobserved intersections, which confirms our intuition. To establish when a vehicle gains or loses privacy, the current state of the vehicle is partitioned into distinct sets of events. At any discrete time t, a vehicle is either in an observed area, or an unobserved area, and within an unobserved area it is either changing pseudonyms, crossing an unobserved intersection or just driving. Thus, four sets of events can be defined, namely T_{upc} for all samples when a vehicle changes pseudonyms unobserved, T_{ui} when a vehicle crosses an unobserved intersection, T_{urs} when it drives on an unobserved road, and finally T_{obs} when the vehicle is observed by a sniffing station. The resulting hybrid privacy loss can be seen in Equation 1.

$$P_{\mathsf{pnm}}(t) = \begin{cases} max(P_{\mathsf{pnm}}(t-1) - \sum_{i=1}^{N_{\mathsf{veh}}} p_i \cdot logp_i, P_{\mathsf{pmax}}) & \text{if } t \in T_{\mathsf{upc}} \\ 0 & \text{if } t \in T_{\mathsf{obs}} \end{cases}$$

$$P_{\mathsf{pnm}}(t) = \begin{cases} max(P_{\mathsf{int}}(t-1) - \sum_{j=1}^{N_{\mathsf{road}}} p_j \cdot logp_j, P_{\mathsf{rmax}}) & \text{if } t \in T_{\mathsf{ui}} \\ 0 & \text{if } t \in T_{\mathsf{obs}} \end{cases}$$

$$P_{\mathsf{pnm}}(t) = \begin{cases} max(P_{\mathsf{road}}(t-1) + \lambda(t_{\mathsf{last}} - t), P_{\mathsf{dmax}}) & \text{if } t \in T_{\mathsf{urs}} \\ 0 & \text{if } t \in T_{\mathsf{obs}} \end{cases}$$

$$P(t) = P_{\mathsf{pnm}}(t) + P_{\mathsf{int}}(t) + P_{\mathsf{road}}(t) \quad (1)$$

This gives $P(t)$, a measure of privacy of the pseudonym used by a vehicle at time t. Privacy is gained according to the sum of the three sources of privacy. The first way a vehicle can gain privacy is similar to the privacy loss function given by [4] and is when an unobserved pseudonym change occurs at time t ($t \in T_{\mathsf{upc}}$). This is given by $P_{\mathsf{pnm}}(t)$, and the new level of pseudonym privacy is then the previous level, modified by the amount of entropy gained by the pseudonym change. The amount of entropy gain is dependent on N_{veh}, the number of vehicles that collaborated with the pseudonym change, and p_i, the probability that each vehicle is the target vehicle. A vehicle can also gain privacy by crossing an unobserved intersection at time t ($t \in T_{\mathsf{ui}}$). The level of intersection privacy is given by $P_{\mathsf{int}}(t)$, and is the previous level modified by the amount of entropy gained by crossing the intersection. The amount of entropy gain is dependent on N_{road}, the number of roads the vehicle can take at the intersection and p_j, the probability of taking each road. Note that these two events can happen at the same time, and both sources of entropy will be added to $P(t)$, the total privacy level at time t. The final way to gain privacy is due to the uncertainty of the exact location of the vehicle when it drives on an unobserved road section at time t ($t \in T_{\mathsf{urs}}$). The level of privacy gained is dependent on t_{last}, which is the last time the vehicle was observed. Thus the longer the vehicle has driven since the last observation, the more privacy is gained. How fast the level of privacy increases is dependent on the sensitivity parameter λ, which models how well the attacker can predict the vehicle's exact position on any given road section. The total level of privacy $P(t)$ is then the sum of these three privacy sources when the vehicle is not observed at time t ($t \notin T_{\mathsf{obs}}$), and a vehicle loses all location privacy when it is observed ($t \in T_{\mathsf{obs}}$), as the attacker now knows the location of the observed pseudonym.

All the methods to increase privacy are limited by a maximum value. In the case of pseudonym changes, the maximum level of gained entropy is limited by the number of vehicles in the tracking area, as it is not possible for an attacker to be confused between more vehicles than are present. This maximum can then be given by $P_{\mathsf{pmax}} = log(N_{\mathsf{tv}})$, where N_{tv} is the total number of vehicles that an attacker can be confused by in the tracking domain (i.e. the campus in the following experiment). The same is true for the entropy gained by crossing unobserved intersections. In this case, the attacker cannot confuse the road an attacker is on between more than the total number of roads in the tracking domain and is given by $P_{\mathsf{rmax}} = log(N_{\mathsf{tr}})$, where N_{tr} is the total number roads in the tracking domain. Finally, the uncertainty that an attacker has in the exact location on a road of a vehicle cannot exceed the length of the longest road, P_{dmax}. The range of values for $P(t)$ is then from a minimum of 0 to a maximum of $P_{\mathsf{pmax}} + P_{\mathsf{rmax}} + P_{\mathsf{dmax}}$.

5.2 Example

Figure 2 illustrates the privacy level of a vehicle over time. In this Figure, a short period of time from the experiment described in Section 6 is considered, where $N_{\mathsf{road}} = 35$ and it is assumed there are 65 vehicles in the area, giving $N_{\mathsf{veh}} = 65$. Furthermore, the attacker considered does not care about where on a road section a vehicle is, giving $P_{\mathsf{road}} = 0$. This then gives a maximum privacy level of $P_{\mathsf{pmax}} + P_{\mathsf{rmax}} + P_{\mathsf{dmax}} = log_2(35) + log_2(65) + 0 = 11.15$. Privacy gains from pseudonym changes are shown by dashed lines and gains from intersections are shown by solid lines. It can be seen that at $t = 0$ the privacy level is 0. A short time after this, the vehicle starts to gain privacy, through a combination of pseudonym changes every 5 minutes, and from crossing unobserved intersections. At $t = 5$, the maximum privacy level from crossing intersections (P_{int}) has reached its maximum (P_{rmax}), and the vehicle can only gain privacy by changing pseudonyms. This happens twice between $t = 5$ and $t = 13.5$, and at $t = 13.5$ the vehicle attains the maximum level of privacy achievable in this scenario. Finally, at $t = 17$ the vehicle comes within range of a sniffing station, and the privacy level is reduced back to 0.

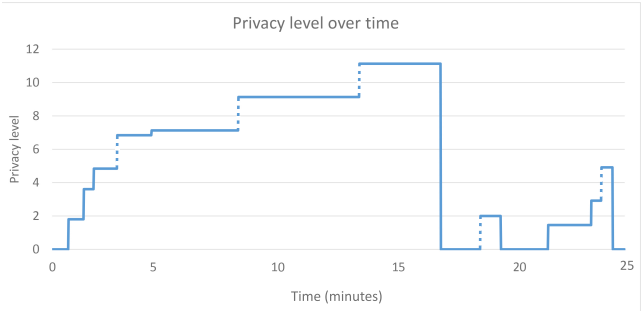

Figure 2: Output of privacy loss function over time

We consider that the privacy level of a vehicle drops to 0 as soon as it is observed, a worst case scenario that highlights the benefit of pseudonyms. In reality, this may however not necessarily be the case. For example, receiving a single CAM from a vehicle at an intersection may give an attacker no information on its ingress or egress direction, and thus, will not completely reduce the privacy level to 0. This decrease in privacy would be better modelled by a decay function. By relaxing this assumption, the benefit of pseudonyms would be even more significant.

6. EXPERIMENT

6.1 Setup

To empirically investigate the effects of vehicle-to-x communications on privacy, two different types of hardware were deployed. Firstly, a transmitting station was installed in a vehicle, which would transmit messages that an attacker can eavesdrop. Secondly, sniffing stations were deployed to actually eavesdrop on these messages and use these to track the vehicle. For the sniffing stations the Cohda Wireless MK3 platform was used. With its built-in 802.11p radio, the Cohda platform allows two antennas to be connected for 802.11p connectivity. High-gain Smarteq V09/54 antennas with a 9 dBi gain were used in combination with the platform. As transmitting station for the vehicle, a Nexcom VTC6201 was used, expanded with a Unex CM10-HI Mini-PCI module with custom drivers for 802.11p connectivity. This module allows for the connection of two antennas and it

Figure 3: The battery, battery charger and Nexcom in-vehicle computer

has an SMA connector for a GPS module. For antennas, two MobileMark ECOM9-5500 were used, covering a frequency range of 5.0-6.0 GHz. These are high-gain 9dBi antennas, with a magnetic mount so that they can easily be fixed to the roof of a vehicle. The Nexcom device was powered by the 12V connector of the vehicle. However, this meant that as soon as the vehicle turned off, the power to the computer would be cut and it abruptly turned off as well. To prevent this from happening, a battery buffer and a battery charger were added. The complete set-up can be seen in Figure 3. The in-vehicle computer can be seen on the right hand side, whereas the battery charger and battery itself can be seen on the left. All equipment was screwed onto a mounting board which could be placed securely in the trunk of the security guard vehicle.

The sniffing stations were deployed at intersections, as such locations maximize the number of vehicles that come within range [5, 7, 14] and offer a wide unobstructed view of the roads connecting to the intersection from different directions. An additional advantage of observing intersections is that the turn-off that a vehicle takes at an intersection in large part determines its route until it reaches the next intersection, where it can turn again. Of course, if there are unobserved intersections in between, then there is always the chance that the vehicle takes a turn onto a different road before the next observed intersection is reached, but it does allow some inferences. For example, if an attacker can observe two non-consecutive intersections, and a vehicle is observed in range of the first intersection and then in range of the second intersection within some time frame, then it can reasonably concluded which road the vehicle took without needing complete reception coverage over this road. This means that an MA can attempt to track a vehicle by using its limited coverage and street-level knowledge of the road network between coverage areas.

6.2 Graph-based View of the Road Network

An MA by definition has some limitations to her resources and the number of sniffing stations she can deploy. This means that an MA also needs to choose which intersections to cover, and which to leave as uncovered gaps. To determine which intersections on the University of Twente campus may give the most information that can be used to track a vehicle, key intersections and interconnecting roads were represented as a graph. Intersections are represented by vertices in this graph, and interconnecting roads as edges. The resulting graph for the campus can be seen in Figure 4.

Figure 4: Turning road network into a graph

The graph gives an abstracted view of the road network, and not all roads and intersections are included. For example, when there are two routes between two adjacent intersections (and there are no roads leading from these roads to other intersections), then this is represented by a single edge. Therefore, the graph gives a high-level overview of which intersections are connected to each other, without being concerned with the smaller details of the actual road network. An intersection covered by a sniffing station can be represented by removing the corresponding vertex from the graph. The remaining graph then represents where a vehicle can travel freely without being in range of an attacker.

This graph was utilized to determine which intersections to cover, and three criteria were defined to help with this. Firstly, an intersection with a large number of connecting roads gives information on all of these roads. With the speed and bearing of a vehicle, an attacker can know exactly which road a vehicle came from and is going to, allowing an attacker to infer a vehicle's position over this larger number of roads. In the graph, this translates to vertices with a large degree (i.e. an intersection with a large number of connecting roads), giving more information than vertices with a low degree. Therefore, sniffing station placement should focus on vertices with largest degree in the graph.

The second criterion are the so-called articulation points of a graph. An articulation point in a graph is a vertex that when removed will completely partition the graph into different biconnected components. This is useful for an attacker, because if this vertex is covered, then the attacker will always know in which biconnected component of the graph a vehicle is. In other words, there would be no route that a vehicle could take from one biconnected component to another

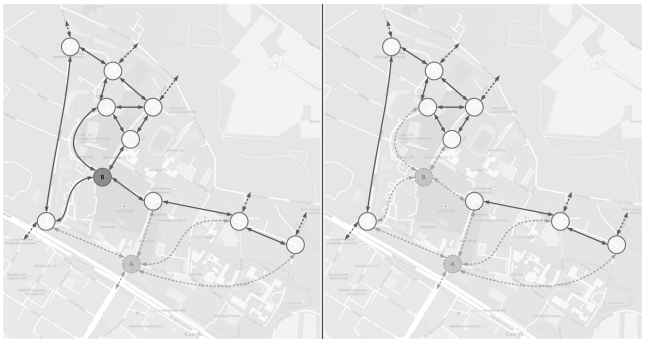

Figure 5: Intersection graph after covering (a) vertex A and (b) vertices A and B

biconnected component, without passing through the intersection that the attacker observes. This allows an attacker to narrow down the position of a vehicle to a certain section of the area within which it wants to track this vehicle.

The third criterion is to cover the busiest intersections, as vehicles are more likely to pass by these. This results in a larger total time that a vehicle is in range of the sniffing station, and thus leaks more information that can be used to track it. An attacker could gain this information by looking at historical traffic data, or even by first employing a learning phase where statistical traffic data is gathered.

Using these criteria, we determined which intersections were good candidates for sniffing stations placement on the campus. Looking at the graph, it can be seen that there is a single vertex that has a degree larger than the others, namely the vertex labelled 'A' which has a degree of 5. This is also the main entrance to the university and thus a busy intersection, complying to the third criterion as well. Therefore this intersection was chosen as the location to place one of our sniffing stations. For the placement of the second sniffing station, articulation points were identified. The vertex labelled 'B' was not an articulation point in and of itself. However, having decided that intersection A would be covered, the corresponding vertex could be removed from the graph. This gave the situation as shown in Figure 5(a), where intersection B becomes an articulation point. This meant that covering both intersections A and B split the entire graph into two different biconnected components where vehicles could travel without crossing an intersection with a sniffing station, as shown in Figure 5(b). Furthermore, these two biconnected components consisted of the residential part of the campus (the western biconnected component) and the university part of the campus (the eastern biconnected component). Hence, vehicles could not travel from one zone to the other without being observed by a sniffing station, giving the attacker insights which could compromise privacy.

Having decided which intersections to cover, one should then decide where exactly at the intersections to place them. To give as good a coverage of the intersections as possible, the sniffing stations needed to be placed close to the intersections. They also needed to be placed somewhere where they were protected from the elements and preferably with an internet connection to allow for remote log retrieval and checking the operational status. This limited the placement to buildings that were near the relevant intersections. For the selected intersections this led to a simple choice, as there was only one building near each of the intersections that could be used. For intersection A this gave a distance of approximately 75 meters between the sniffing station and the centre of the intersection. For intersection B the distance between the centre of the intersection and the sniffing station was approximately 110 meters.

6.3 Data Collection

Two sniffing stations were deployed to determine vehicle trackability with a minimal number of sniffing stations. The transmitting station in the vehicle and the sniffing stations were deployed for a total of 16 full days. During this time, approximately 300MB of CAM data were collected on all stations combined. The vehicle logged all transmitted CAMs, representing the ground truth. The sniffing stations logged all eavesdropped CAMs, representing the observed data. The vehicle took 411 trips, and transmitted 2,734,691 CAMs in a total time of approximately 76 hours. The logs from the sniffing stations on the other hand contained just over 68,542 eavesdropped CAMs. This meant that the sniffing stations managed to pick up messages from the vehicle for a total time of approximately 1.9 hours, and that only 2.5% of all transmitted messages were eavesdropped. On average the vehicle drove for 4.75 hours per day, of which 7.1 minutes within range of a sniffing station.

The log files were processed to remove GPS errors and periods where the vehicle was stationary for prolonged periods of time. Cleaning up the vehicle's log removed 53.56% of CAMs transmitted, going down to 1,270,016 CAMs. This represented about 38.24 hours of useful driving data. On the sniffing stations, 40,254 CAMs remained after cleaning, a reduction of 41.27%. Of these remaining messages, 18,293 were received at intersection A at the main entrance of the university and 21961 were received at intersection B. Our eavesdropped messages then consisted of 3.17% of all transmitted CAMs, and covered about 1.1 hours of vehicle driving time. Whilst this seems low, one should note that the eavesdropped time is not equal to the tracking time, as an attacker will attempt to infer where the vehicle was in the periods where no messages were eavesdropped. Section 7 and Appendices A and B show that a considerable level of tracking can be achieved even with this small percentage of eavesdropped messages.

7. RESULTS

7.1 Application of Privacy Loss Function

The privacy loss function was applied to evaluate the level of privacy that the vehicle had in the experimental scenario. If the vehicle did not use pseudonyms, this meant that $P_{\mathsf{pmax}} = 0$ and the vehicle could not gain any privacy through pseudonym changes alone. However, the vehicle could still gain some privacy through crossing unobserved intersections. The maximum level of intersection privacy that an attacker could obtain was $P_{\mathsf{rmax}} = log(N_{\mathsf{tr}}) = log(35)$, as there were 35 roads in the tracking area. The level of entropy that a vehicle gained by passing an unobserved intersection depends on the number of roads leaving an intersection, and the probability that the vehicle took each road. For example, an intersection where a vehicle almost always takes the

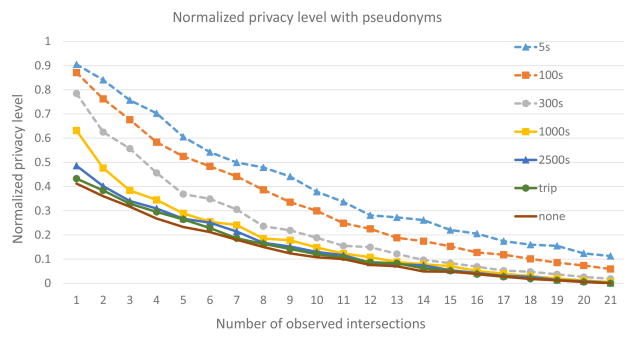

Figure 6: Privacy level for different pseudonym change strategies

same road will give less entropy than an intersection where the vehicle is equally likely to take each road. This data was already available for the vehicle in the ground truth, and this was used to establish the probabilities of each road being taken. For each intersection ingress event, how often each road at that intersection was taken was calculated. This gave the probability p_j in the loss function, and allowed calculation of the entropy gained for each unobserved intersection that the vehicle could pass.

To compare different pseudonym change strategies, the overall privacy level that a certain change strategy results in needed to be determined. To do this, the mean privacy level over all location samples was calculated. The normalized privacy level of the vehicle in our experimental scenario can be seen in Figure 6. This figure shows the privacy level of the vehicle for different numbers of observed intersections. This takes into account two different pseudonym change strategies, namely per trip and periodic. With pseudonyms, a vehicle can gain entropy as the attacker may confuse the target vehicle with all other vehicles that change pseudonyms on the same road at the same time. The factor that influences this privacy level is the number of vehicles that change pseudonyms at the same time, and so for each change $p_i = \frac{1}{N_{\text{veh}}}$. Furthermore, the total amount of privacy gained by pseudonyms is limited by the total number of vehicles present. To analyse pseudonyms in our experimental scenario, assumptions need to be made about these variables, based on what was considered to be realistic for the test site. Firstly, we assume a 'perfect pseudonym change' (i.e. change where there is enough generated uncertainty that an attacker cannot link the old pseudonym and the new pseudonym, and hence cannot conclude that they belong to the same vehicle), thus it is assumed that every time a pseudonym was changed, there were three other vehicles that changed pseudonyms at the same time. Furthermore, it was assumed that there were 100 vehicles in the tracking area, giving $P_{\text{pmax}} = log(N_{\text{tv}}) = log(100)$. With these assumptions, the effect of pseudonyms on the mean level of privacy in our experimental scenario can be seen in Figure 6.

The level of privacy decreases as the number of observed intersections increases, and changing pseudonyms more often gives a higher level of privacy. Furthermore, changing pseudonyms more often seems to be most effective when there are only a few observed intersection. When the number of observed intersections increases, the pseudonym change strategy used seems to have less influence on the privacy level. In fact, for a pseudonym change period of 2500 seconds, the privacy level is just marginally better than not using pseudonyms at all. This is due to the fact that with many observed intersections, the chance is high that a vehicle will come across such an intersection quickly, and all gained privacy will be lost. Thus, with more observed intersections, the larger the influence of intersections on the privacy level, and with fewer observed intersections the more pseudonym changes affect the privacy level. For short pseudonym periods a higher privacy level is obtained, even with many observed intersections. These results are however slightly biased. Due to the computational complexity of calculating the average privacy level for all combinations of all number observed intersections, a random selection of intersections was taken. Hence, this does not reflect exactly the selection that an attacker might make, as a smart selection of intersections might decrease the privacy level even further.

This graph can be used to get an indication of what the effects of pseudonyms are on the resource level of the attacker. For example, to get the same privacy level, the attacker needs to cover only 1 intersection in the case of no pseudonyms, but needs to cover 10 intersections with a pseudonym change period of 5 seconds. Therefore, introducing pseudonyms increases the resources that an attacker needs to commit to be able to track a vehicle to the same extent. Furthermore, the mean level of privacy does not tell the entire story. Intersections cover a geographic area in the tracking domain, and thus, there will be some areas with higher level of privacy than others. Intuitively, areas with higher density of observed intersection will result in a lower privacy level, but also the probability of which roads are taken will affect this level. To get a visual indication of privacy, a privacy heatmap was generated. The privacy heatmap of our experimental scenario with two observed intersections, and a pseudonym change period of 300 seconds, can be seen in Figure 7. In this Figure, it is clear that there is an overall low level of privacy around the observer intersections. Only around unobserved intersection is the privacy level somewhat higher, with privacy being the highest the furthest away from the observed intersections. By using a heatmap to visualize our hybrid privacy loss function, it is possible to get an intuitive understanding of how the placement and quantity of sniffing stations affects the privacy level of a vehicle. For more results regarding zone-level tracking and road-level tracking, we refer the reader to Appendices A and B.

7.2 Cost Model

The attacker resources determine the number of intersections that can be observed. These resources consist of time to install hardware, computational resources, knowledge to analyze eavesdropped messages, and financial resources to purchase sniffing stations. Installing sniffing stations and obtaining the computational resources and knowledge needed to track a vehicle is easy. The main limiting factor is the financial costs to purchase sniffing stations.

The cheapest solution for a sniffing station including antennas came to approximately €500. To cover all intersections on the campus of the University of Twente, an attacker

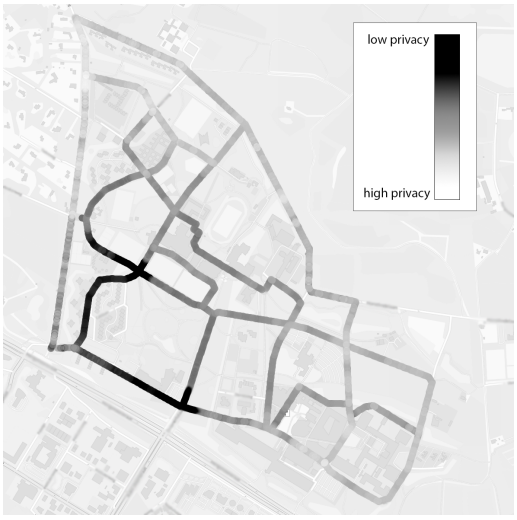

Figure 7: Privacy heatmap for an attacker covering two intersections (A and B)

would need to place a sniffing station at 21 intersections. This would mean a financial cost of 21× €500 = €10,500. However, not all intersections may need to be covered to track a vehicle, and adding more intersection results in diminishing returns. Experimental results show that only 8 of the 21 intersections are sufficient to track a vehicle on a road-level up to 90% of the time. An attacker would then only require €4000. If the attacker is interested in zone-level tracking, only two intersections on zone boundaries need to be observed to correctly predict whether the vehicle was in the residential or university part of the University of Twente 78% of the time. If five intersections are observed on the zone boundaries, this prediction accuracy increases to more than 95% (see Table 1 of Appendix A). One should note that for zone-level tracking, the attacker does not need an accurate location. Nevertheless, on a road-level tracking, we managed to get accurate location with an error of maximum 20 meters. The scale can also be expanded by looking at the area that an attacker can cover. The total area of the University of Twente was approximately 1.75 km^2. Assuming the 21 sniffing stations required to cover this intersection, this would mean that an attacker would require approximately 12 sniffing stations per 1 km^2 that she wants to cover. Given a sniffing station cost of €500, this results in €6,000 per km^2. If this is extrapolated to the entire city of Enschede with a total area of approximately 143 km^2, an attacker would require 1716 sniffing stations, and an attacker would need €858,000 to fully cover the entire city. This is however only if the attacker does not consider the road network and assumes that all intersections need to be covered. If an attacker does not require 100% coverage and uses the road network, significantly less financial resources would be required. For example, if it is assumed that an attacker only needs to observe 8 out of every 21 intersections, then the costs of covering the whole city already drops to €327,000.

One should note that we assume a sniffing station price of €500, because as the technology is relatively new, a large part of these costs cover research and development of the hardware. It is expected that prices will drop significantly in the upcoming years, as more competition comes to market, and the production quantity increases to match increased demand from ITSs being deployed. Beyond this, it is likely that in the future other devices will be able to receive ITS messages as well. Especially considering that 802.11p is just a modification of 802.11a, we do see an influx of bringing 802.11p functionality to mobile phones [16]. Another option would be to use an inexpensive computer such as a Raspberry Pi with an 802.11a dongle and a driver patch to support 802.11p. This would already bring the costs of a single sniffing station down to approximately €50. Following the example above of an attacker that wants to cover the city of Enschede using road network knowledge, using this hardware would bring the total costs down to €32,700. The similarity of 802.11p to 802.11a also opens up other attack vectors. For example, if an attacker can compromise a large number of 802.11a routers found in homes and patch them to receive 802.11p messages as well, it could be possible to quickly create a sniffing station botnet covering a considerable geographic area. Finally, the US Department of Transportation has announced intention to mandate OBU in all new vehicles in the near future. This means that it would be possible to obtain sniffing stations from junk yard.

How pseudonyms affect the costs for an attacker can now be considered. The main effect that pseudonyms have on the financial resources required by an attacker, is that they increase the number of sniffing stations that an attacker needs to deploy to get the same tracking capabilities as without pseudonyms. Looking at the normalized privacy level in Figure 6, using no pseudonyms and covering 8 intersections leads to a normalized privacy level of 0.16. To get the same privacy level with a vehicle that uses pseudonyms and changes every 300 seconds, the attacker would need to cover 12 intersections, an increase of 50%.

The use of pseudonyms consistently makes it more difficult for an attacker to track a vehicle to the same level as when no pseudonyms are used, and so it is always advisable to use a pseudonym scheme. Apart from increasing the number of sniffing stations required to track a vehicle to the same level, using pseudonyms may also force an attacker to change her tracking methods (e.g. when a desired level of tracking can never be reached, even by covering all intersections).

8. CONCLUSION AND FUTURE WORK

Privacy is a cornerstone of successful adoption of ITS by the general public because connected vehicles will broadcast location beacons, enabling location tracking. This is a desirable feature at a local level (i.e. below one kilometer) as this cooperative awareness is key to safety applications. However, a larger trackability is not required for safety applications or traffic efficiency, and thus, should be prevented by privacy protection mechanisms. Numerous pseudonym schemes have been proposed in the past, but none were seriously considered by standardization bodies yet. Perhaps this is due to the general misconception that location privacy attacks are only of interest when considering a global attacker with extensive resources. So, to encourage standardization bodies and governments to foster research regarding this issue, we demonstrated that location tracking is accessible to anyone with at least two 802.11p-capable devices.

In this paper we presented results from the first real world experiment focused on tracking capability of a mid-sized observer and pseudonym change frequencies. Experiment results demonstrate that location tracking is easy to perform, and that two sniffing stations are sufficient to offer 40% road-level tracking, while eight sniffing stations offer 90%. We also introduced a cost model that gives an overview of the attacker resources (i.e. number of sniffing stations she can deploy) with respect to the privacy level for different pseudonym change frequencies. This helps regulators, policy makers and implementers to identify the appropriate pseudonym change frequency in function of the assumed attacker resources. Results also demonstrated that even if pseudonyms cannot completely prevent location tracking attack, they can significantly mitigate the risk.

We are now extending the tracking domain to city-wide scenario to analyze the impact of the tracking domain scale and its intersection density. These results will refine our cost model. We have already found that there is no significant difference between a campus-wide and city-wide road network such as Orlando. However, we have noticed that road networks following a Manhattan grid style provide an overall better privacy level. This is mostly because the topology makes timing attack (i.e. correlation between ingress and egress times used to decrease the anonymity set size) difficult. This brings the idea of 'privacy enhancing road networks', where road networks are designed with the concept of privacy at their core. We will also propose to use 'cities-level privacy' to adjust privacy protection (e.g. pseudonym change strategy). Indeed, if the system knows the driver is going to/through a city with overall low privacy level, then it could increase the pseudonym change frequency or enable more sophisticated mechanisms.

References

[1] ETSI TS 102 637-2 V1.2.1 - Intelligent Transport Systems (ITS);Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service. Technical report, ETSI.

[2] L. Buttyán, T. Holczer, and I. Vajda. On the effectiveness of changing pseudonyms to provide location privacy in VANETs. *Lecture Notes in Computer Science*, 4572:129, 2007.

[3] M. Duckham and L. Kulik. Location privacy and location-aware computing. *Dynamic & mobile GIS: investigating change in space and time 3*, 2006.

[4] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. C. Parkes. On non-cooperative location privacy: A Game-Theoretic Analysis. In *16th ACM conference on Computer and communications security (CCS)*, pages 324–337, 2009.

[5] J. Freudiger, R. Shokri, and J.-P. Hubaux. On the optimal placement of mix zones. In *Privacy enhancing technologies*, pages 216–234. Springer, 2009.

[6] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5(4):38–46, 2006.

[7] M. Humbert, M. H. Manshaei, J. Freudiger, and J.-P. Hubaux. Tracking games in mobile networks. In *Decision and Game Theory for Security*, pages 38–57. Springer, 2010.

[8] J. Krumm. Inference Attacks on Location Tracks. *Pervasive Computing*, 10(Pervasive):127–143, 2007.

[9] S. Lefevre, J. Petit, R. Bajcsy, C. Laugier, and F. Kargl. Impact of V2X privacy strategies on Intersection Collision Avoidance systems. *5th IEEE Vehicular Networking Conference (VNC)*, pages 71–78, Dec. 2013.

[10] J. Petit, M. Feiri, and F. Kargl. Revisiting attacker model for smart vehicles. In *IEEE 6th International Symposium on Wireless Vehicular Communications (WiVeC)*, pages 1–5, Sept 2014.

[11] J. Petit, F. Schaub, M. Feiri, and F. Kargl. Pseudonym schemes in vehicular networks: A survey. *IEEE Communications Surveys Tutorials*, PP(99):1–32, Sept 2014.

[12] SAE International. SAE J2735 V1.1.1 - Dedicated Short Range Communications (DSRC) Message Set Dictionary. Standard, 2009.

[13] R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. Le Boudec. Quantifying location privacy: The case of sporadic location exposure. In *Privacy Enhancing Technologies*, volume 6794 of *Lecture Notes in Computer Science*, pages 57–76. 2011.

[14] O. Trullols, M. Fiore, C. Casetti, C. Chiasserini, and J. Barcelo Ordinas. Planning roadside infrastructure for information dissemination in intelligent transportation systems. *Computer Communications*, 33(4):432–442, Mar. 2010.

[15] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos. Privacy in inter-vehicular networks: Why simple pseudonym change is not enough. In *7th International Conference on Wireless On-demand Network Systems and Services (WONS)*, pages 176–183. IEEE, Feb. 2010.

[16] X. Wu. Future Technology Trends for Vehicular Communication. *Keynote International Symposium on Wireless Vehicular Communications (WiVec)*, 2014.

[17] B. Yu, C.-Z. Xu, and B. Xiao. Detecting Sybil attacks in VANETs. *Journal of Parallel and Distributed Computing*, 73(6):746–756, June 2013.

APPENDIX

A. ZONE-LEVEL TRACKING

To investigate the relationship between the number of observed intersections and the extent of possible zone-level tracking, an attacker that had the resources to cover more than two intersections was considered. By observing the ingress and egress events at observed intersections, it may be possible to determine in which region the vehicle was at a specific time. In order to emulate such a scenario, we assumed that the attacker could observe (at least) a 35 meter radius around the centre of an observed intersection. Given the fact that a sniffing station could also be weatherproofed, an attacker could place a station close to an intersection even

Figure 8: Intersections on zone boundaries

when there are no buildings nearby. Such a sniffing station could be battery powered, as it was found that the sniffing stations could run for more than 24 hours using the battery shown in Figure 3. Thus, with a practical signal range of around 300 meters, a 35 m coverage area around an intersection is a conservative assumption, and an attacker would be able to cover most intersections in this manner. The CAMs from within the 35 meter area were then added to the set of eavesdropped messages. The information from these messages was sufficient to determine the speed and trajectory of the vehicle crossing an intersection, and so identify ingress and egress events and directions.

The next step was to determine how observed intersection events translated to zone predictions. Both ingress and egress events at the two intersections were taken into account. If an ingress or egress event was observed, the direction of the vehicle was used to determine what zone the vehicle was in. For egress events it was assumed that the vehicle stayed within this zone until the next observed event. For ingress events, it was assumed the vehicle was in the corresponding zone since the last observed event. However, if the vehicle traveled from one zone to the other unobserved, then it was possible to record an egress event into one zone, and then some time later observe an ingress event from another zone, giving conflicting information. To solve this, the time between these observations was divided into two equal parts, and the egress zone was assigned to the first part and the subsequent ingress zone to the second.

Five intersections were identified on the zone boundary between the residential zone and the university zone, as shown in Figure 8. The next step was to determine the prediction accuracy if the attacker had observed them. The prediction accuracy of all different combinations of observed intersections was calculated, where the zone was predicted according to observed intersection events, and then compared to the actual zone as given by the ground truth. The prediction accuracy was then the percentage of correctly predicted location samples. The prediction accuracy for each number and each combination of observed intersections can be seen in Table 1. For example, row "1-3-5" means that the attacker covered 3 intersections (1, 3, and 5), which gave a prediction accuracy of 77.44%.

As expected, covering only a single intersection gave the worst prediction accuracy, whereas covering all 5 identified intersections leads to the best prediction accuracy of 95.28%. The remaining incorrect predictions could be attributed to zone transitions where the vehicle did not have a GPS fix, as this meant that the sniffing station could not eavesdrop on the vehicle's trajectory and identify ingress and egress events. For each additional sniffing station, the average prediction accuracy increases by approximately 8.5%. This could be useful for an attacker to determine the trade-off between the costs of an additional sniffing station and the desired prediction accuracy. From these results we can conclude that, given sufficient resources, an attacker can collect the information required to accurately predict a vehicle's most likely zone.

B. ROAD-LEVEL TRACKING

To determine the effect of the number of observed intersections on road-level tracking, all large intersections within the tracking domain were identified. Figure 10 shows the 21 large intersections found, along with their connecting routes.

Note that even when all intersections in a tracking domain are observed, the observer is still a mid-sized attacker and not a global attacker which could eavesdrop on all messages in the entire area; due to the limited radio range, the areas between intersections are considered unobserved. As in Appendix A, it was assumed that an attacker could observe the 35 meters surrounding the centre of each intersection. The attacker could then fully track a vehicle within this intersection area, but additionally could infer which roads connecting to the intersection the vehicle was on. More specifically, when observing an egress event, the attacker can infer the vehicle location up to the next intersection on this road. After this point the vehicle has the possibility to change roads, and if this intersection is not observed, then the vehicle can no longer be tracked. By observing an intersection, the road a vehicle is on is fully known until the vehicle has the opportunity to change its route at the next unobserved intersection. Vice versa, when observing an ingress event, an attacker could infer the vehicle's past location up to the last intersection that it crossed.

With these assumptions, the percentage of all CAMs where the attacker knew either exactly where the vehicle was (when it was in range of a sniffing station), or exactly on which road section the vehicle was (by inference as described above), was calculated. This was done for every number of covered intersections between 1 and the maximum of 21. Furthermore, every combination of intersections that an attacker could cover was considered. This gave a total of $\sum_{i=0}^{21} \binom{21}{i} = 2097152$ different combinations of intersections that an attacker of various resources could cover. The results of these calculations can be seen in Figure 9. This Figure shows both the maximum and mean tracking percentage for all combinations of n intersections. The maximum tracking percentage is the maximum that an attacker can obtain with n intersection, out of all possible combinations

# of intersections	1		2		3		4		5	
	1	61.12%	1-2	72.82%	1-2-3	81.40%	1-2-3-4	84.26%	1-2-3-4-5	95.28%
	2	67.49%	1-3	73.42%	1-2-4	78.96%	1-2-3-5	89.51%		
	3	54.85%	1-4	67.41%	1-2-5	81.53%	1-2-4-5	86.41%		
	4	52.53%	1-5	69.98%	1-3-4	73.15%	1-3-4-5	86.58%		
	5	58.10%	2-3	73.32%	1-3-5	77.44%	2-3-4-5	87.29%		
			2-4	71.76%	1-4-5	74.33%				
			2-5	78.62%	2-3-4	77.38%				
			3-4	61.44%	2-3-5	83.74%				
			3-5	67.66%	2-4-5	82.09%				
			4-5	59.10%	3-4-5	72.50%				
average		58.82%		69.55%		78.25%		86.81%		95.28%

Table 1: Zone-level prediction accuracy for all intersection combinations

of this number intersections. The combination of observed intersections that leads to this maximum is the optimal combination. The mean tracking percentage is the mean out of all these combinations, with the error bars indicating one standard deviation error. The maximum tracking performance quickly increases as more intersections are observed. For example, to achieve a tracking rate of 90%, only 8 intersections need to be covered. The optimal combination for 8 intersections which leads to the maximum tracking percentage was intersections 3, 4, 8, 10, 14, 17, 18 and 21. The situation when these 8 intersections are observed is shown in Figure 10, where the black routes and intersections represent areas where we can fully track a vehicle on a road-level, and the grey routes and intersections are where the vehicle can move freely without being tracked.

In this configuration there are a number of isolated grey intersections. These are intersections where the attacker can infer the vehicle's location on all roads connecting to the intersection, but the intersection itself is not observed. However, as the roads extend to the centre of the intersection, these isolated grey points do not indicate an area where the vehicle is safe from tracking. Apart from these isolated intersections, we can see that there are 8 road sections where the vehicle cannot be tracked on a road-level. Zone-level tracking is still possible however, as an attacker can infer that when the vehicle enters a grey zone that it remains in this zone until it is observed again. Moreover, within these zones the road a vehicle is on might still be inferred by using the timing attack. However, even without timings there is still some information that an attacker can gain on unobserved zones. For example, if a vehicle is observed leaving intersection 17 towards intersection 20, and some time later is observed entering intersection 18 from the south, then the vehicle must have taken the grey route between intersection 19 and 20, as there was no other unobserved way to get there. In fact, the only time when an attacker will not know what road, or set of connecting roads, the vehicle is on is when there is a loop in a grey zone, such as between intersections 5, 6 and 7. As all the roads in this loop are approximately of equal length, a timing attack is not possible, and the road network actually facilitates a vehicle's privacy level.

Finally, it may also not be of interest to an attacker to be able to track a vehicle on the road-level inside the entire tracking domain. For some areas in the tracking domain, zone-level knowledge of a vehicle may suffice, whereas in other areas road-level tracking is required. For example, a burglar in the residential zone will be interested in the exact road the security vehicle is on in this zone, but will not care about where the exact road when it is in the university area, just that it is in this zone. Thus, an attacker could create a denser network of sniffing stations where more tracking information is needed, and for other areas where only zone-level tracking is necessary, the sniffing stations would only need to be placed on the zone boundaries.

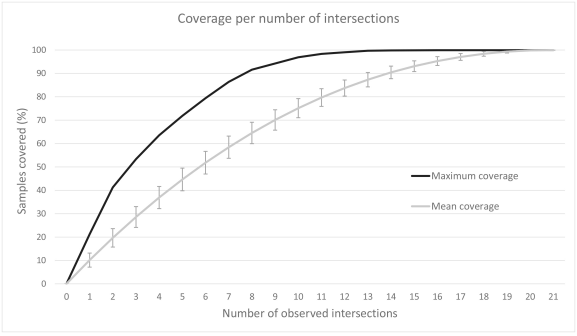

Figure 9: Road-level tracking percentage for all intersection combinations

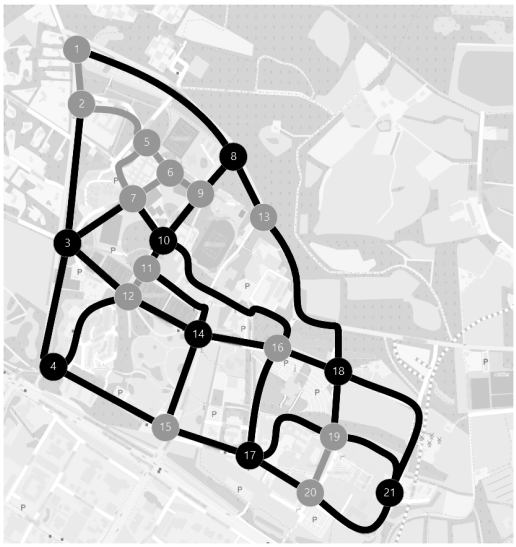

Figure 10: Optimal coverage for 8 intersections