TECHNISCHE UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR INFORMATIK

Lehrstuhl für Angewandte Informatik - Kooperative Systeme

Safety-aware Location Privacy in Vehicular Ad-hoc Networks

Karim Ahmed Awad El-Sayed Emara

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr.rer.nat.)

genehmigten Dissertation.

Vorsitzende:	Univ.-Prof. Dr. Claudia Eckert
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Johann Schlichter 2. Univ.-Prof. Dr. Uwe Baumgarten

Die Dissertation wurde am 11.11.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 13.02.2016 angenommen.

Gedruckt mit Unterstützung des Deutschen Akademischen Austauschdienstes

Acknowledgment

All praise and thanks to Allah, who provided me the ability to complete this work. I would like also to thank all people who helped me make it possible. This thesis would not have been possible without the help and support of my supervisors, colleagues and family.

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Johann Schlichter, for his continuous support, patience, motivation, time and guidance. He always cared about my progress and he often gave me sufficient time to thoroughly discuss the challenges encountered during my research. I would also like to thank my second supervisor, Prof. Dr. Uwe Baumgarten, for his insightful comments and support, especially in writing recommendation letters annually for extending my scholarship. I would also like to thank Dr. Wolfgang Wörndl for his close support and discussion during my research.

I want to thank all my colleagues at the chair of Applied Informatics and Cooperative System. Among them, Dr. Michele Brocco, Dr. Georg Groh, Hubert Kreuzpointner, Alexander Lehmann and Dr. Benno Schweiger are notable. I always benefit from their comments and feedback, especially during Ph.D. colloquiums. I appreciate discussion with other peers in the community during scientific events. I also thank anonymous reviewers of my papers for their valuable and constructive feedback to improve my work. I would like to especially thank Bjoern Wiedersheim, Ulm University for providing me their MHT tracker and STRAW vehicle traces.

Grateful acknowledgements should be also delivered to the German Academic Exchange Service (DAAD), Egyptian Ministry of Higher Education and TUM Graduate School for their financial support.

I would never finish this work without the support of my parents and family. I thank them for always supporting and praying for me. Their love and encouragement have been and will always be a great source of inspiration in my life. Finally, my special thanks go to my lovely wife for her support and patience to fulfill my career goals.

Abstract

Vehicular Ad-hoc Networks (VANET) provide wireless communication among vehicles to exchange information for better traffic safety and efficiency. Safety applications broadcast beacon messages periodically and unencryptedly which contain a pseudonym, a time stamp and the vehicle state (position, speed and heading). Pseudonyms are changed regularly to avoid messages linkability. However, beacons of the old and new pseudonyms are still linkable by exploiting their spatiotemporal information. If the adversary is global and covers a sufficiently large area of the road network, it could track all vehicle movements. Furthermore, the adversary can identify the drivers' sensitive whereabouts, social activities and personal preferences remotely and globally without control or knowledge of the driver. This privacy risk must be handled to ensure the public acceptance of VANET.

Although there are some privacy schemes for VANET, only few schemes consider their impact on safety applications. Privacy schemes are usually composed of anonymization along with data obfuscation or beacon elimination. These mechanisms reduce the quality of the exchanged information and may hinder the operations of safety applications. Therefore, it is essential to analyze the impact of privacy schemes on safety applications, when designing or evaluating a privacy scheme.

In this dissertation, we focus on preserving location privacy without hindering the operations of safety applications. To accomplish this goal, we investigated methods of measuring both the location privacy and quality of service (QoS) of safety applications. To measure the location privacy, a robust and efficient vehicle tracker was developed that achieves a high tracking accuracy with vehicle traces of various densities, position noises and beaconing rates. This tracker acts as a global adversary which we employed to measure the protection level of a privacy scheme. Using this tracker, typical location privacy metrics were also discussed and compared. Moreover, we adapted a practical and extensible methodology based on Monte Carlo analysis to measure the QoS of two safety applications, forward collision warning and lane change warning. This methodology is applicable to any privacy scheme and can be extended to measure the QoS of other applications.

We proposed and evaluated obfuscation privacy schemes showing their ineffectiveness in preserving privacy and their significant negative impact on

safety applications. Also, two context-aware privacy schemes were proposed that consider both the vehicle context and driver preferences to determine the appropriate situations to change pseudonyms. In addition, we provided a quantitative and qualitative comparison between our proposed schemes and other privacy schemes proposed in literature. We employed both simulated and realistic vehicle traces in all evaluations which provides high trustworthiness in the presented results.

The experiment results show that it is possible to preserve location privacy with small impact on the QoS of safety applications. In general, location privacy is not preserved by only frequently changing pseudonyms (even synchronously), but a discontinuity in the spatiotemporal information is additionally required to prevent tracking. A best compromise is to remain silent for a short period synchronously and globally among all vehicles before a pseudonym change. A practical compromise between privacy and QoS is to select the appropriate context where a vehicle should change its pseudonym and remain silent. Also, choosing the appropriate privacy metric is essential because non-representative metrics results in overestimation of the preserved privacy. A metric based on the distortion between the tracks that are reconstructed by an adversary and the actual traces is effective to measure the privacy level. Moreover, QoS metrics should reflect the ability of safety applications to calculate their requirements rather than estimating the expected distance error or delay in communication.

Privacy And Safety
Location Privacy
Vehicular Network
Vanet
Beacons
Qos Metric
Distortion
Mix Zone
Privacy Metric
Tracker
Silent Period
Quality Of Service
Pseudonym
Privacy Preference
Context Adaptive
Context Aware
Security
Pseudonym Change
Impact On Safety
Security And Privacy
Privacy Scheme
Traceability
Vehicle State
Safety Applications
Vehicle Traces
V2x Communication
Car2x Communication
Forward Collision Warning
Lane Change Warning
Tracker Confusion
Global Adversary
Qos Metric
Privacy And Safety
Vanet
Beacons
Security

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AS	Anonymity Set
ASS	Anonymity Set Size
BSA	Basic set of applications
BSM	Basic Safety Message (aka beacon message)
CA	Certification Authority
CADS	Context-Adaptive Privacy Scheme
CAM	Cooperative Awareness Message (aka beacon message)
CAPS	Context-Aware Privacy Scheme
CCW	Cooperative Collision Warning
CMIX	Cryptographic MIX-zone privacy scheme [53]
CPN	Cooperative Pseudonym based on number of Neighbors [97]
CRL	Certificate Revocation List
CSP	Coordinated Silent Period
DGPS	Differential Global Positioning System
DLR	Deutschen Zentrums für Luft- und Raumfahrt (German Aerospace Center)
DoS	Denial of Service
DOT	Department of Transportation in United States
DSRC	Dedicated Short Range Communication
ETSI	European Telecommunications Standards Institute
FCC	U.S. Federal Communications Commission
FCW	Forward Collision Warning
GNN	Global Nearest Neighbor (Data Association method)
GPA	Global Passive Adversary
GPS	Global Positioning System
IBC	Identity-based Cryptography
ICA	Intersection Collision Avoidance
IMA	Intersection Movement Assist
ITS	Intelligent Transportation System
JPDA	Joint Probabilistic Data Association
JPDAF	Joint Probabilistic Data Association Filter

LAA	Local Active Adversary
LBS	Location-based Service
LCW	Lane Change Warning
MAC	Medium Access Control (communication layer) OR Message Authentication Code
MANET	Mobile Ad hoc Network
MHT	Multi-Hypothesis Tracking (Data Association method)
MTP	Mean Tracking Percentage
MTT	Multiple Target Tracking
NHTSA	National Highway Traffic Safety Administration
NNPDA	Nearest Neighbor Probabilistic Data Association
OBU	On-board Unit
OV	Other vehicle
PDA	Probabilistic Data Association
PDR	Packet delivery ratio
PKI	Public Key Infrastructure
QoS	Quality of Service
REP	Random Encryption Period [138]
RSP	Random Silent Period privacy scheme [71]
RSU	Road Side Unit
RWP	Random Way Point (mobility model)
SAE	Society of American Engineers
SLOW	A context-based privacy scheme [28]
STRAW	Street Random Way Point (mobility model)
SUMO	Simulation of Urban Mobility (traffic simulator)
SV	Subject Vehicle (The equipped or concerned vehicle)
TAPAS	TAPAS Cologne scenario is one of the largest freely available traffic simulation data set based on the SUMO traffic simulation
TPD	Tamper Proof Device
TTC	Time to Collision
VANET	Vehicular Ad hoc Network
V2I	Vehicle-to-Infrastructure communication
V2V	Vehicle-to-Vehicle communication
V2X	Vehicle-to-Anything (vehicle, infrastructure, human, etc.)
VISSIM	Verkehr In Städten - SIMulationsmodell (Traffic in cities - simulation model)
VSC	Vehicle Safety Communication (consortium)
WAVE	Wireless Access in Vehicular Environments (IEEE 1609)
WSMP	WAVE Short Message Protocol

List of Publications

The following list is the author's publications closely related to this thesis:

Refereed Journals:

1. **Karim Emara**, Wolfgang Woerndl, and Johann Schlichter, "Context-based Privacy Schemes for VANET," *EAI Endorsed Transactions on Security and Safety*, Invited paper. (submitted)

2. **Karim Emara**, Wolfgang Woerndl, and Johann Schlichter, "On Evaluation of Location Privacy Preserving Schemes for VANET Safety Applications," *Computer Communications*, 63:11-23, June 2015. (**IF: 1.695**)

Refereed Conferences/Workshops:

3. **Karim Emara**, Wolfgang Woerndl, and Johann Schlichter, "POSTER: Context-Adaptive User-Centric Privacy Scheme for VANET," *11th EAI International Conference on Security and Privacy in Communication Networks (SecureComm)*, Dallas, USA, October 2015.

4. **Karim Emara**, Wolfgang Woerndl, and Johann Schlichter, "CAPS: Context-Aware Privacy Scheme for VANET Safety Applications," *ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)*, pp. 21:1-21:12, New York, USA, June 2015. (**AR: 19%**)

5. **Karim Emara**, Wolfgang Woerndl, and Johann Schlichter, "Vehicle tracking using vehicular network beacons," in *4th International Workshop on Data Security and PrivAcy in wireless Networks (D-SPAN)*, IEEE WoWMoM, Madrid, Spain, Jun. 2013.

6. **Karim Emara**, "Location privacy in vehicular networks," in *5th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks PhD forum 2013 (IEEE WoWMoM 2013 PhD forum)*, Madrid, Spain, Jun. 2013.

Technical Report:

7. **Karim Emara**, Wolfgang Woerndl, and Johann Schlichter, “Beacon-based Vehicle Tracking in Vehicular Ad-hoc Networks,", Technical Report in *Technical University of Munich, Department of Informatics*, April 2013.

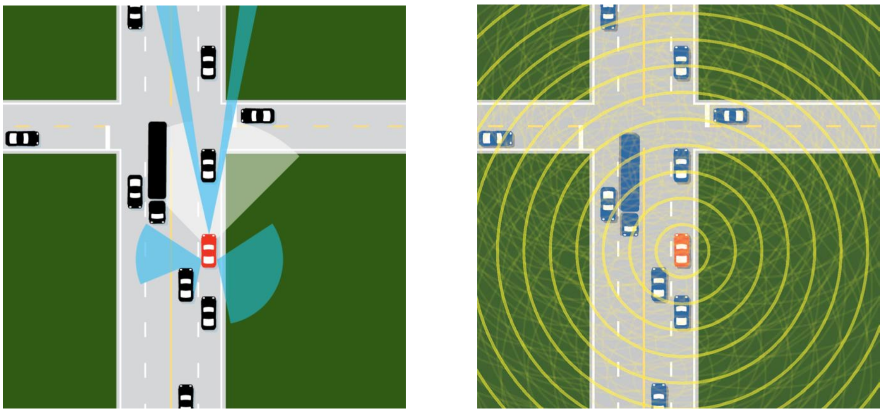
1 Introduction

Vehicular adhoc networks (VANET) provide wireless communication among vehicles to exchange information autonomously. Over the last decade, VANET has gained considerable interest in both research and industry for safety, traffic efficiency and infotainment applications. It is evident that VANET will be realized in near future to minimize traffic fatalities and support self-driving cars. Our goal in this dissertation is to preserve location privacy in VANET without significantly reducing the quality of service (QoS) of safety applications. This goal is attained by proposing privacy schemes and evaluating them with respect to their privacy protection level and their impact on the QoS.

In this chapter, we will present our motivation, objectives and research questions. A brief overview of the research methodology, system models and vehicle traces are also explained.

1.1 Motivation

Connected and cooperative vehicles are mandatory for future intelligent transportation system. The benefits of vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) communication are numerous and involve wide areas of safety and traffic efficiency. There are several large-scale field operational tests that have been conducted already in Europe (simTD in Germany [3], DRIVE C2X in Europe [2]) and in U.S. (Safety Pilot Program [1]) which confirmed the effectiveness of VANET applications in reducing crashes. According to the analysis conducted by the U.S. Department of Transportation's (DOT) National Highway Traffic Safety Administration (NHTSA), crashes, injuries, and fatalities could be reduced by 50% on average using two potential safety applications, intersection movement assist and left turn assist [64]. A fully mature VANET system of V2V and V2I communication could potentially address 81% of all vehicle target crashes involving unimpaired drivers [92]. These safety benefits let the U.S. DOT accelerate its timetable on the proposed VANET rule that would require V2V equipment in all new vehicles [51, 93].



(a) Non-cooperative (b) V2V communication

Figure 1.1: V2V communication offers more comprehensive awareness than other detection systems (e.g., radar, camera) [Source: [64]]

To attain the benefits of safe and efficient traffic, VANET applications broadcast *beacon*[1] messages periodically and publicly. A beacon message usually contains the vehicle state (position, speed and heading) along with a pseudonym which is changed periodically according to a pseudonym-change scheme [105]. These beacons enable a 360-degree awareness of surrounding vehicle states and possible threats, as illustrated in Figure 1.1. The information exchanged among vehicles is so precise in position and time to be able to support the requirements of safety applications. Shladover and Tan [122] claim that a positioning accuracy up to 1 m is required for most cooperative collision warning applications.

Since this information is broadcast unencryptedly, a serious privacy threat arises if all these beacons are collected and analyzed. Although pseudonyms are changed periodically, beacons of the old and new pseudonyms are still linkable by exploiting the spatiotemporal information in beacons [27]. If the VANET adversary would cover a sufficiently large area of the road network, it could track all vehicles remotely and continuously. Having an external adversary who can cover the whole network seems very difficult, but we assume the worst case scenario. In addition, this model could be realizable through an untrusted service provider and its deployed roadside units. Moreover, the accuracy and frequency of VANET beacons are much higher than those expected from other systems such as traffic monitoring cameras and location-based ser-

[1] Beacon message is also known as Basic Safety Message (BSM) in U.S. standards and Cooperative Awareness Message (CAM) in European standards.

vices which make the revealed private information about users more accurate and detailed.

Although the exchanged beacons are anonymous and contain no identifying information, further privacy attacks can be performed. The important observation about vehicle traces is that they are almost unique on their own in most of the cases. The start and end points and times, the frequency over week and month and the routes followed are highly discriminating features for vehicle traces. For example, it is rare to find two neighbors who go to work in the same or near places every day at the same time and follow the exact route. Also, quasi-identifiers such as vehicle attributes (e.g., size, type) that would be included in beacons can differentiate among mixed traces. Moreover, driving characteristics, whether originating from the vehicle capabilities or driver behavior, can be exploited to identify traces of a vehicle, similar to work done in [151]. Based on these features, the de-anonymization of anonymous traces is achievable using work/home pairs [59] or top N locations [152] and with the help of geosocial networks [33]. Once the traces are de-anonymized, the adversary can identify the user's sensitive whereabouts, social activities and personal preferences remotely and globally without control or knowledge of the user. These privacy risks must be handled to ensure the public acceptance of VANET.

Although there are some schemes that handle continuous tracking in VANET, only few schemes consider their impact on safety applications. Privacy schemes are usually composed of anonymization along with data obfuscation or elimination [124]. These mechanisms reduce the quality of the exchanged information and may hinder the operations of safety applications. For example, if the privacy scheme decided to keep silent at a safety critical situation, it could prevent the safety application to produce a timely alert. Therefore, it is important to analyze the impact on the quality of service (QoS) of safety applications, when designing or evaluating a privacy scheme. The trade-off between privacy and safety is sporadically studied in literature and still considered as an open research and deployment challenge, according to a recent survey of Petit *et al.* [105].

1.2 Objectives and Research Questions

In this thesis, we aim to protect vehicles from continuous tracking (through beacon messages) without hindering the operations of safety applications. This goal is divided into the following objectives:

O1. Develop a state-of-the-art vehicle tracker which is able to track exchanged beacons effectively in different traffic conditions.

O2. Investigate the existing location privacy metrics and propose a suitable metric for the VANET scenarios.

O3. Measure and assess the impact of privacy schemes on the quality of service of safety applications.

O4. Design and evaluate one or more location privacy schemes which prevent vehicle tracking with a minimal impact on safety applications.

O5. Compare the proposed privacy schemes with the existing state-of-the-art schemes in terms of privacy and safety levels.

According to the presented objectives, this thesis tries to answer the following **research questions**:

RQ1. Based on the fact that vehicle movements are predictable, what is the most suitable and efficient tracking algorithm for the VANET beaconing use case? To what extent can beacons be tracked compared with other tracking methods?

RQ2. What are the main factors that prevent beacon tracking the most? (e.g., position accuracy, beaconing rate, type of information)

RQ3. How to measure the location privacy? And what is the most suitable metric for the VANET beaconing use case that ensures correctness, generality and practicality?

RQ4. How to measure the impact of privacy schemes on safety applications? Given the diversity of safety applications, is it possible to provide a generic measurement methodology that is applicable to different applications?

RQ5. Is it effective and safe to use the obfuscation schemes to preserve location privacy in VANET?

RQ6. How efficient are the context-based privacy schemes? Do they offer a better compromise between privacy and safety?

RQ7. Based on the proposed privacy and safety metrics, how effective are the existing schemes in preserving privacy compared with the proposed schemes? To what extent do they affect the operations of safety applications?

1.3 Contributions

The main contributions of this thesis are:

1. Develop a robust vehicle tracker that achieves a high tracking accuracy with vehicle traces of various densities, position noises and beaconing rates. It outperforms the commonly-used multi-hypothesis tracker (MHT) in tracking accuracy and efficiency.

2. Propose embedding a vehicle tracker inside vehicles which increases the awareness of the vehicle about the surrounding traffic even if neighbor beacons are missed or noised.

3. Adapt a practical and extensible methodology for measuring the impact of a privacy scheme on two safety applications, forward collision warning and lane change warning.

4. Propose and evaluate obfuscation schemes showing their ineffectiveness in preserving privacy and their negative impact on safety applications.

5. Propose two context-aware privacy schemes that consider both the vehicle context and driver preferences to determine the appropriate situation to change pseudonym.

6. Evaluate privacy schemes using robust tracker and realistic vehicle traces in terms of a representative privacy metric. This reflects the credibility of the presented results when compared with other related works.

1.4 Methodology

In this thesis, a quantitative simulation-based approach is adopted to measure both privacy and safety levels which facilitate attaining our goal. As stated above, we aim to preserving location privacy in VANET without hindering the operations of safety applications significantly. We employ various vehicle traces of different mobility models to evaluate and compare different privacy schemes. We look at vehicle traces as if they are broadcast by vehicles in a fully penetrated VANET and collected by a global passive adversary. Vehicle traces are then modified according to the specifications of the privacy scheme such as changing the pseudonym and eliminating some beacons during silent periods.

Location privacy is quantified by measuring how effective a tracker can reconstruct vehicle traces from the collected beacons. For this reason, a robust vehicle tracker, based on a multi-target tracking technique, is developed and

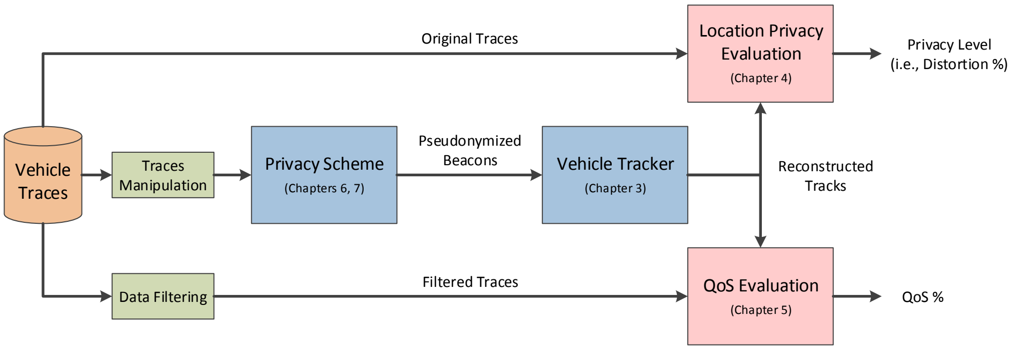


Figure 1.2: Overview of the system blocks

evaluated, as discussed later in Chapter 3. The reconstructed traces by the tracker are compared with the original vehicle traces to calculate the distortion percentage which expresses on the privacy level, as will be described in Chapter 4. In addition, the QoS of safety applications is evaluated by estimating the probability of correctly identifying the fundamental requirements of a safety application using Monte Carlo analysis. Two safety applications are considered which are forward collision warning (FCW) and lane change warning (LCW) applications. We choose these applications because they require the most precise location information (<1 m) and the highest beaconing rate (10 Hz) [38]. More details about the QoS evaluation can be found in Chapter 5. Moreover, several privacy schemes are proposed, evaluated and compared with existing privacy schemes, as will be explained in Chapters 6 and 7.

Figure 1.2 illustrates these building blocks and how they interact. Starting from vehicle traces, they are obtained from a traffic simulator or a realistic traces dataset, as will be explained in Section 1.6. They generally consist of vehicle ID, position and velocity in xy coordinates. They are manipulated so that they look like beacons broadcast from vehicles by adding noise or dropping some packets. A Gaussian noise of 50 cm standard deviation is added to the position of each coordinate. In some experiments, random beacons are eliminated every time step to simulate the effect of packet loss. The considered privacy scheme modifies the manipulated beacons by adding pseudonyms and changing them according to the scheme specifications. It may also obfuscate or eliminate beacons to simulate the effect of obfuscation or silence periods. The pseudonymous beacons obtained from the privacy scheme are given to the vehicle tracker to be reconstructed into tracks. The reconstructed tracks are compared with the original traces to calculate the distortion percentage and with the filtered traces to obtain the QoS of safety applications. Given the unified distortion and QoS percentages, we can flexibly compare different privacy schemes with respect to their compromise between privacy and safety levels.

Positioning

The research work on security and privacy in VANET can be structured horizontally according to the underlying authentication technique and vertically according to the pseudonym life cycle. The authentication technique can be based on the public-key infrastructure, group signature, identity-based cryptography or symmetric cryptography. The pseudonym life cycle includes issuance, usage, change, resolution, and revocation phases. On the one hand, this thesis considers the pseudonym change phase in the public-key infrastructure. On the other hand, the impact of privacy schemes on the safety applications is also considered and measured.

1.5 Technical Models

In this section, we describe the system and adversary models which will be adopted in subsequent chapters.

1.5.1 System Model

We assume each vehicle is equipped with an on board unit (OBU) which it uses to communicate with other vehicles and broadcast *beacon* messages periodically (1-10 Hz). The beacon information includes a pseudonym, a timestamp and the current vehicle state (i.e., position, speed and heading). Vehicles use a state-of-the-art pseudonym issuing process such as [78] to retrieve a pool of pseudonyms to be used one by one in the V2X communication. Pseudonyms have a *minimum pseudonym time* during which they must be kept unchanged to ensure stable communication. After that time, a vehicle changes the pseudonym according to the adopted privacy scheme. The European standard ETSI TS 102 867 recommends changing a pseudonym every five minutes [8] while the American SAE J2735 standard recommends changing it every 120 s or 1 km, whichever comes last [6]. Since beacons are essentially used by safety applications, the broadcast information has to be as precise as possible. Thus, we assume each vehicle is equipped with a GPS receiver and combines the obtained measurements with its internal sensors to minimize the position error up to 50 cm. This small error is recommended in [122] and also realized in systems such as [120] to be able to achieve useful Cooperative Collision Warning applications (CCW). We assume that a vehicle maintains the states of its nearby vehicles located within its communication range (e.g., 300 m) using a multi-target tracking (MTT) algorithm. The utilization of a MTT algorithm for neighbor states maintenance is two-fold. First, it allows a vehicle to predict, with the help of a Kalman filter, the state of neighbors even if their beacons are delayed or missed due to a communication error or a silence period. As a result, the MTT algorithm can enhance the effectiveness of safety applications. Second, the MTT algorithm supports the vehicle in choosing the appropriate situation to change pseudonyms that increases the likelihood of tracker confusion.

1.5.2 Adversary Model

We assume a global passive adversary (GPA) that deploys low-cost receivers over a large part of the road network and eavesdrops on all exchanged messages. Having an external adversary that can cover the whole network may seem challenging, but we assume the worst case scenario. Also, this model is realizable, for example, by an untrusted service provider through its deployed roadside units. The main objective of the adversary is a *tracking attack* or reconstructing all vehicle traces from their beacon messages. Thus, we assume that the driver's location privacy is determined by the protection level against this attack. Although breaching the driver's location privacy requires de-anonymization of the reconstructed traces, the de-anonymization process is out of the thesis scope. However, we assume that the more complete and correct the reconstructed traces, the more successful the de-anonymization process.

The adversary achieves its objective by correlating the beacons of a vehicle by pseudonym matching. When a vehicle changes its pseudonym, the adversary uses a multi-target tracking algorithm to correlate the messages of the old and new pseudonyms. If the adversary covers only a small part of the road network, it can still track vehicles within this limited area, but such tracking may not be valuable regarding de-anonymization as it does not reflect complete traces. Although powerful adversaries can track vehicles using already-deployed cameras spread over the road network, the cost and inefficiency of global camera-based attacks will be much higher than those for global beacon-based attacks [53].

1.6 Vehicle Traces

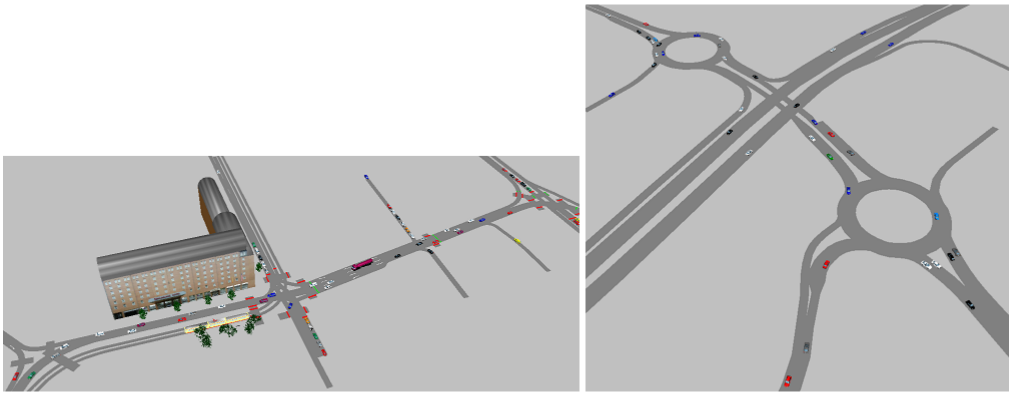
We use several vehicle traces datasets in evaluating different parts of this thesis. In general, we use traces generated from traffic simulators in measuring the effect of different parameters on the measured entity. Additionally, we use realistic traces to verify its applicability in real-world situations. We use traces generated from the VISSIM simulator [60] and the STRAW (STreetRAndom Waypoint) vehicular mobility model [36]. The realistic vehicle traces are obtained from [135]. Next, we explain the details of each dataset.

1.6.1 VISSIM Traces

The VISSIM simulator is a microscopic and behavior-based simulation that models the vehicle traffic and public transport operations. It uses a microscopic traffic flow simulation model including the car following model and lane change logic [108]. The VISSIM uses a psycho-physical driver behavior model developed in [142]. The basic concept of this model is that the driver of a faster vehicle starts to decelerate as she reaches her individual perception threshold to a slower vehicle in front. Since she cannot exactly determine the speed of that vehicle, her speed will fall below that vehicle speed until she starts to slightly accelerate again after reaching another perception threshold. This behavior results in an iterative process of acceleration and deceleration. The VISSIM supports also significant control on the road network and traffic customization. It supports drawing roads and connection links between them, adding priority rules, stop signs and traffic signals. It allows traffic composition of several vehicle types and characteristics. The traffic arrival rate, vehicle desired speed and route decisions can be efficiently configured in the VISSIM graphical interface. The VISSIM also supports logging the vehicle and network information on a discrete time basis down to 100 ms. We used VISSIM for its realistic mobility model and variety of parameters which provides an effective control on the generated traces.

We employed the logging feature to generate vehicle traces every 100 ms. The trace file includes the position in the three coordinates, scalar values of speed and acceleration, along with the vehicle ID. The vehicle heading is not directly generated from VISSIM, therefore we calculated it using positions of every two consecutive time steps. Finally, the velocity and acceleration vectors are calculated for each coordinate. Thus, the final vehicle traces consist of the position, velocity and acceleration in the three coordinates along with the vehicle ID and grouped per time step.

We choose two scenarios included in the VISSIM demos that represent urban and highway road networks. The urban scenario is a part of roads in Luxembourg city and consists of three main intersections controlled by fixed-time traffic signals, and five join and exit roads, as shown in Figure 1.3(a). The main road is multi-lane single direction and is crossed by two-direction single-lane roads. The total length of all roads is about 3.18 km. The Figure 1.3(b) shows the highway scenario which consists of a multi-lane two-direction main road



(a) Urban Network

(b) Highway Network

Figure 1.3: The main parts of the road networks of the VISSIM scenarios

with two roundabouts and a bridge with total road length of 3.87 km. As this network represents a highway, there is no traffic signal or stop sign.

For both scenarios, the simulation duration is 300 s which is sufficient for traffic to enter and exit the network several times with all different routes. The routes that vehicles follow are pre-configured in the VISSIM network files. However, we changed the density and distribution of the traffic by changing the arrival rates at all entry points and the vehicle desired speed. There is an entry point located in the start point of each road and one can control the vehicle arrival rate at each point. Since it is important to evaluate the impact of traffic density, we generated several datasets with different arrival rates. We selected the ranges of 100 - 600 and 300 - 1000 vehicle/hour in the urban and highway scenarios, respectively. These ranges result in a maximum number of simultaneous vehicles of 25 - 195 and 20 - 64 vehicles in the urban and highway scenarios, respectively. These arrival rates are chosen to avoid frequent long traffic jams. We also generated several datasets for different vehicle desired speeds. The desired speed is that the driver seeks during the simulation and tries to keep when there is nothing hindering the vehicle. Thus, it is not necessary for vehicles to drive in such speed constantly; their actual speed depends on the surrounding traffic and the logic of the mobility model. The desired speeds are assigned to vehicles randomly based on the configured distribution. We assigned a uniform distribution of desired speeds around the specified value. We selected the desired speeds of 30 - 70 km/h and 80 - 130 km/h in the urban and highway scenarios, respectively. When the desired speed is varied, the default arrival rate of 300 and 600 vehicle/hour is used in the urban and highway scenarios, respectively. Similarly, when the arrival rate is varied, the default

Table 1.1: Parameters of urban and highway scenarios in VISSIM traces

Parameter	Scenario	Range	Default value
Arrival Rate (Vehicle/h)	Urban	100 - 600	300
	Highway	300 - 1000	600
Desired Speed (km/h)	Urban	30 - 70	50
	Highway	80 - 130	100
Max Simultaneous Vehicles	Urban	25 - 195	77
	Highway	20 - 64	35
Total Roads Length (km)	Urban		3.18
	Highway		3.87
Sampling Interval (s)	Both		0.1
Simulation Time (s)	Both		300

desired speed of 50 km/h and 100 km/h is used in the urban and highway scenarios, respectively. The last parameter is the sampling interval which is assigned to 0.1 s, because many safety applications require an update frequency of 10 Hz [38]. These parameters are summarized in Table 1.1.

1.6.2 STRAW Traces

The STRAW traces are generated by Wiedersheim et al. [143]. They have a road map of 1 km^2 and are generated from the STreetRAndom Waypoint (STRAW) mobility model [36] on Central Boston map for 1000 s. It provides accurate simulation results compared with the Random Waypoint (RWP) mobility model because it uses a vehicular mobility model of real cities in the United States, based on the operation of real vehicular traffic [36]. As described in [143], the STRAW model simulates vehicle movements in traffic networks that are composed of road segments, which are sub-divided into lanes. The number of traffic signals, the number of lanes in each direction, and the maximum speed differ on the basis of the street type. The vehicles in each lane periodically calculate the acceleration or deceleration for the next time step. Because no collision recognition is implemented, vehicles that simultaneously cross an intersection may collide. Vehicles cannot change lanes in the road segment, except when entering a new road segment. The vehicle density is kept constant in each trace file by making vehicles route within road segments and never exit.

The original traces contain the vehicle ID, time stamp, and position in a 1 s stepping. We calculate the velocity assuming a constant velocity between every

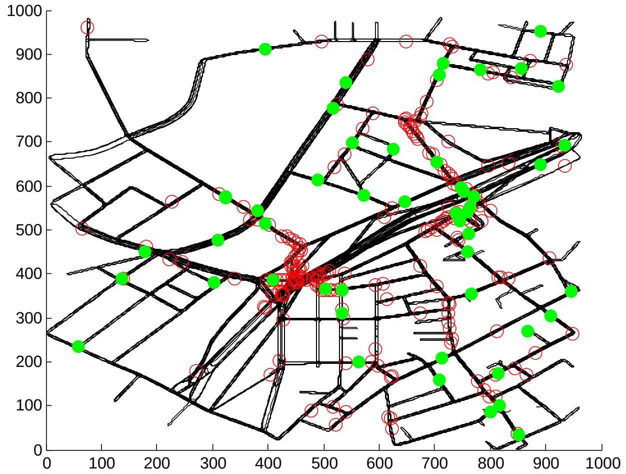
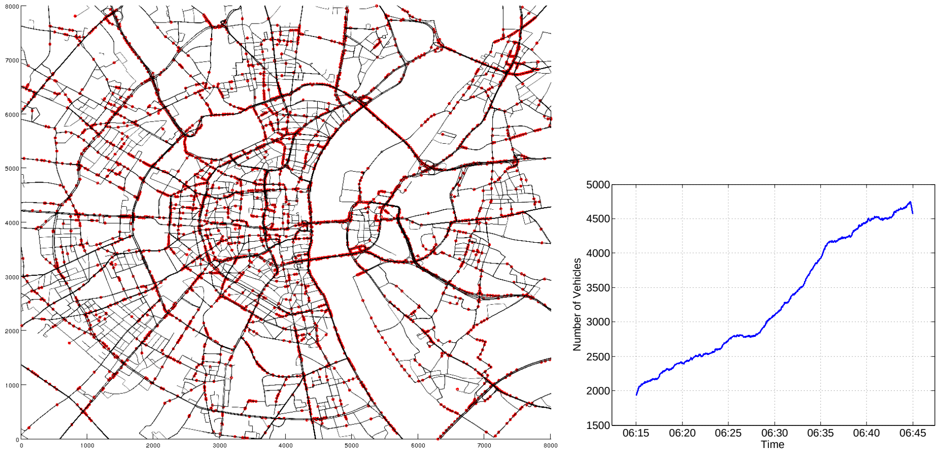


Figure 1.4: The road map of the STRAW traces

two consecutive time steps and interpolate the samples to create a 0.5 s stepping. The maximum vehicle speed ranges from 11 to 26 m/s depending on the road, the maximum acceleration is 2.23 m/s^2 and the maximum deceleration is 11.15 m/s^2. The road map of these traces is shown in Figure 1.4 where snapshots of the sparsest case of 50 vehicles and densest case of 200 vehicles are represented by green points and red circles, respectively. Each vehicle density has 10 variations with different routes.

1.6.3 Realistic Traces

The realistic vehicle traces are obtained from [135]. This dataset is mainly based on the data made available by the TAPASCologne project [7] which is an initiative by the Institute of Transportation Systems at the German Aerospace Center (ITS-DLR). This dataset reproduces vehicle traffic in the greater urban area of the city of Cologne, Germany with the highest level of realism possible. The street layout of the Cologne urban area is obtained from the OpenStreetMap (OSM) database. The microscopic mobility of vehicles is simulated using the Simulation of Urban Mobility (SUMO). The source and destination of vehicle traces are derived through the Travel and Activity PAtterns Simulation (TAPAS) methodology. Uppoor *et al.* [136] pointed out several problems when combining these data sources to produce traffic data. Among these problems, vehicles are moving rapidly to large traffic jams, travel times are unrealistic and vehicle speeds turn to very low values. Uppoor *et al.* resolved these problems so that the synthetic traffic match that observed in the real world, through real-time traffic information services. This is why we name this dataset as realistic



(a)

(b)

Figure 1.5: (a) The road map of the realistic traces. (b) The vehicle density versus time.

traces.

We obtained the two-hour sample published online [135] and selected 30 min (from 6:15 AM till 6:45 AM) for the middle 64 km^2 region, as shown in Figure 1.5(a). We selected this time period because the vehicle density increases dramatically, which provides a challenging evaluation for the operation of privacy scheme in different densities. As we cropped the vehicle traces in both space and time, we excluded very short traces that move within 100 m^2 or start and end in less than 15 s. There are 19,704 remaining traces with increasing density, ranging from 1,929 to 4,572 simultaneous vehicles in the first and last time steps, respectively, as shown in Figure 1.5(b). The vehicle positions in the last time step are drawn as red spots in Figure 1.5(a). Moreover, we processed the dataset to calculate the heading and velocity in the xy-coordinates using every two consecutive time steps for each vehicle. The last heading value is preserved when the vehicle stops and is changed when it starts to move.

1.7 Thesis Structure

This thesis is structured into eight chapters. **Chapter 1** introduces the whole thesis showing the motivation, objectives and research questions. It also presents the research methodology along with system and adversary models and the employed vehicle traces.

Chapter 2 provides an overview of VANET showing its prospective applications, main characteristics and communication protocols. The security and privacy in VANET is also reviewed discussing their requirements and a detailed survey on the existing approaches.

Vehicle tracking will be discussed in **Chapter 3** which starts with introducing the problem of multiple target tracking along with its necessary components. Then, the developed vehicle tracker is explained showing the underlying motion model used in Kalman filter. The vehicle tracker is evaluated in various traffic densities, position noises, beaconing rates and packet delivery ratios. It is also compared with multi-hypothesis tracker that is typically used in related work.

Chapters 4 and 5 present the adopted metrics for the privacy and QoS of safety applications. **Chapter 4** discusses the existing location privacy metrics and explains the adopted distortion metric. It also presents a comparison among different metrics. **Chapter 5** explains the QoS measurement methodology of safety applications based on vehicle traces modified by a privacy scheme. This methodology is applied on two safety applications: forward collision warning and lane change warning applications.

In **Chapter 6**, two obfuscation privacy schemes are proposed and evaluated: position perturbation and random beaconing rate. They are also evaluated in comparison with random silent period in terms of privacy and safety levels. **Chapter 7** presents the proposed context-based privacy schemes. The first scheme lets vehicles select the effective context in which to enter a silence period, to change its pseudonym and when to resume beaconing with a high probability of confusion to a global adversary. A more advanced scheme is also proposed which adapts its parameters according to the real-time traffic density and the driver's privacy preference. Last but not least, a comparative evaluation among some existing privacy schemes is presented in comparison with the proposed privacy schemes. Finally, **Chapter 8** lists the thesis conclusions and future work.

2 Background

Vehicular ad hoc networks (VANET) have emerged in the past years and gained interest from both academia and industry. Vehicular networks are those networks formed among vehicles (V2V communication) and between vehicles and infrastructure (V2I and I2V communication) to provide diverse traffic-related and infotainment applications. The most important applications of VANET are those aimed at enhancing traffic safety and providing a better driving experience. The principle benefits of VANET include the high quality and quantity of cooperative information among vehicles and infrastructure, the non-line-of-sight knowledge and the potential coordination among vehicles [32]. Although VANET can be envisioned to be the largest realization for mobile ad hoc networks (MANET) serving hundreds of millions of vehicles worldwide [89, 110], protocols and techniques designed for MANET cannot be directly adopted by VANET. In fact, the size of the network, the high speed of vehicles, the sporadic connectivity and the slow deployment process add more challenges to VANET [111].

In VANET, vehicles are supposed to be equipped with computing, sensing, communication and user interface components. The computing platform is dedicated for VANET operations with appropriate interfaces to the in-vehicle system. The on-board sensors are assumed to obtain essential data such as positioning through GPS, velocity, direction, brakes status and airbags status [101]. For the communication components, each vehicle is equipped with an On Board Unit (OBU) which allows one- and multiple hop V2X communications. The OBU connects to the infrastructure through Roadside Units (RSU) installed along the road. Both OBUs and RSUs are supposed to support Dedicated Short Range Communication (DSRC) standard with a bandwidth of 75 MHz in the 5.9 GHz band and a communication range of 100-1000 meters. Over other wireless technologies, DSRC provides significant advantages of very low latency (less than 100 ms) and support for transmitting broadcast messages [38]. In addition, other wireless technologies can be used (such as cellular communication and WiFi) for infrastructural data access and in non latency-critical scenarios.

In this chapter, we provide an intensive introduction to VANET including its applications (Section 2.1), characteristics (Section 2.2) and the underlying wireless technology (Section 2.3). In addition, Section 2.4 reviews security and privacy in VANET explaining their requirements and threat models. Security techniques are discussed in Section 2.5. Last but not least, the privacy schemes are categorized and discussed in Section 2.6.

2.1 VANET Applications

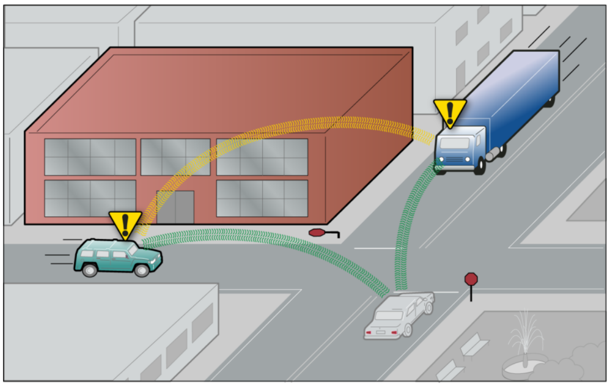
VANET applications can be generally divided into three categories: safety, traffic efficiency and infotainment/other applications. Each application has different requirements and characteristics to operate and fulfill its use cases. These characteristics include communication type (V2V or V2I), transmission mode (periodic or event-triggered), maximum packet size, communication range, minimum message frequency, maximum allowable latency, information accuracy (position accuracy), security level, penetration rate (percentage of vehicles equipped with VANET technology) and the required infrastructure [65, 101].

Effort has been made to identify and evaluate potential application scenarios that should/can be used in the initial deployment phase of VANET. The ETSI presented the basic set of applications (BSA) that can be deployed simultaneously within a three-year time frame after the standards have been completed [5]. These BSAs are selected based on questionnaire results obtained from stakeholders about the societal, customer and business values of the use cases of several applications. Among 75 investigated application scenarios, the Vehicle Safety Communication (VSC) consortium identified eight potential high benefit safety applications whose requirements are assumed to be representative of the requirements for safety applications [38]. Recently, NHTSA reviewed various VANET-based safety applications to verify whether or not VANET could address crashes resulting from the considered circumstances [64]. In addition, many application scenarios and use cases can be found in [13, 47, 80, 132, 144]. Next, we will present application examples from each category.

2.1.1 Safety Applications

Safety applications are those applications that aim at reducing the probability of traffic accidents and consequently saving lives on the road. These applications share information among vehicles and road side units to allow drivers to avoid collisions and hazardous situations [77]. Most of these applications require strict requirements such as low latency of 100 ms, frequent update of 10 Hz and precise vehicle positioning less than 1 m [101].

Forward collision warning (FCW) [120] warns the driver of a likely rear-end collision with a heading vehicle in the same lane and direction of travel. Cur-



Source: GAO.

Figure 2.1: Example of V2V Intersection Movement Assist Warning Scenario. The truck and sports utility vehicle are at risk of colliding because the drivers are unable to see one another approaching the intersection and the stop sign is destroyed. (Source: [64])

rent FCW applications based on visual and radar detection systems cannot operate in poor lighting and weather conditions (sunrise, sunset, rain, snow), and are limited with respect to distance. However, VANET-based FCW applications can function in conditions beyond the visual and radar detection systems [64].

Lane change warning (LCW) [120] warns the driver during a lane change if the blind spot zone, into which the driver intends to switch, is or will be occupied by another vehicle moving in the same direction. The application has the potential to address at least 19% of the crashes in the lane change crash group [64].

Intersection collision warning (ICW) [43, 90] (aka Intersection Movement Assist IMA) warns the driver when it is not safe to enter an intersection due to a high collision probability with other vehicles at controlled (with stoplights) and uncontrolled (with stop, yield, or no signs) intersections. This application might not be available without the VANET technology because it requires awareness beyond the line-of-sight and farther than the range of visual and radar detection systems. The ICW should address five types of junction-crossing crashes which together represent 26% of all vehicle crashes [64]. One example scenario of ICW is illustrated in Figure 2.1.

2.1.2 Traffic Efficiency Applications

Traffic efficiency applications aim at enhancing the efficiency of transportation network by sharing real-time traffic status provided by vehicles, road side units

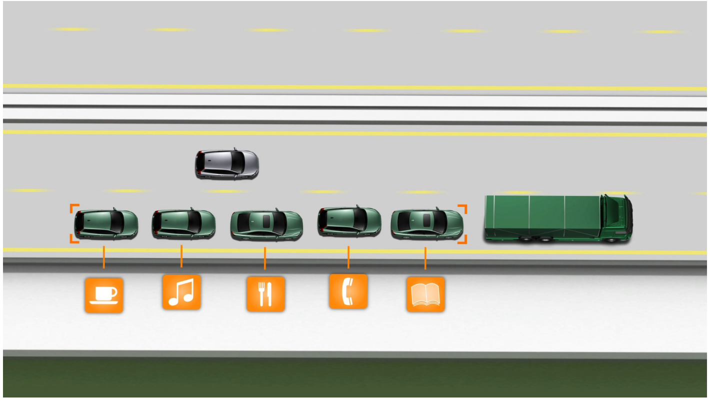


Figure 2.2: SARTRE road trains. Passengers in the five following vehicles can do other activities during platooning. (Source [4])

and other trusted sources. This information may be collected and processed by traffic operators to offer recommendations to drivers to reduce delays and enhance driving experience. Car-to-Car Communication Consortium (C2C-CC) [37] selected several potential use cases such as enhanced route guidance and navigation, green light optimal speed advisory and platooning which are described briefly below.

Enhanced route guidance and navigation [5] uses information collected by an infrastructure administrator about the real-time traffic status to identify congestion, work zones and other factors causing travel delays and to report these delays to navigation systems inside vehicles. Road side units inform vehicles within its region about the current and expected traffic conditions and recommends alternative routes to drivers.

Traffic (Green) light optimal speed advisory [25] provides information to drivers on how to avoid stopping at intersections and traffic lights to make driving smoother and optimize fuel consumption. As a vehicle approaches an intersection, it receives information regarding intersection location and the remaining signal timing (the number of seconds until a red light switches to green). The vehicle can calculate the optimal speed required to reach the traffic light without necessitating stopping or slowing down.

Co-operative vehicle-highway automation system (Platooning) [20] groups vehicles into virtual road trains to increase road capacity by decreasing safe distance needed for human reaction. This application allows many vehicles to accelerate or brake simultaneously following the dynamics of the lead vehicle transmitted over VANET. The challenges with platooning is the coordination of platoon members which is usually done with a platoon leader acting as the controlling vehicle. The benefits of platooning are numerous such as optimized fuel consumption, increased safety, efficient road utilization and better driver convenience. SARTRE project [4] conducted a field test that includes a lead truck followed by three cars driven entirely autonomously at speeds of up to 90 km/h with no more than 6 meters gap between the vehicles, see Figure 2.2.

2.1.3 Infotainment Applications

Infotainment applications do not concern safety or traffic, but rather the drivers' interests and needs. Information and entertainment applications comprise quite a diverse set of scenarios and use cases such as tolling, point-of-interest notifications, fuel consumption management, podcasting and multihop wireless Internet access [65]. More infotainment applications can be found in [38].

2.2 VANET Characteristics

According to the presented applications, it is clear that VANET has unique characteristics when compared with other types of MANET. These characteristics include:

- **High topology change.** Due to the high speed of vehicles, network topology is always changing resulting in sporadic connectivity and difficulty with long session establishment [13, 82]. Also, the content of VANET messages can change the network topology [150]. For example, a driver could be advised by an enhanced route guidance application to change her route to avoid a traffic jam.

- **Large scale and variable density.** VANET can grow to a very large scale especially in city centers and at entrances to big cities [132, 150]. However, vehicle density will be low during the initial deployment resulting in sparse connectivity and network partitioning. In later deployment phases, the density will be related to the location and time. For example, consider a road section of three lanes. In normal cases, 70 vehicles can be found around a given vehicle within 1 km radius assuming 70 m inter-vehicle distance. However, the number of surrounding vehicles might be more than 1000 with 5 m inter-vehicle distance during a traffic jam [132].

- **Predictable mobility.** This is a unique feature in VANET whereby vehicles move in a predefined and known road network. In fact, vehicles are required to follow the road restrictions and rules such as speed limit, direction and traffic lights [74, 82, 132]. However, whereas the predictability of the position of a vehicle allows an improvement in link selection, the linear topology of VANET decreases the possibility of finding a redundant link [132].

- **No significant power and computational constraints.** Since batteries in vehicles are self-rechargeable (at least while driving), power supply is not as critical in VANET as in MANET applications. Vehicles would be equipped with powerful computation resources rather than hand held devices [41, 82].

- **Various communication environments and types.** There are various types of communication in VANET. In highway traffic, the communication environment is straightforward; while it becomes much more complex within cities due to different types of obstacles (e.g., buildings, trees) [82]. Also, vehicular applications often require communication with other vehicles in a specific geographical area or location [131].

- **Built-in positioning capability.** GPS is widely used in modern vehicles for route guidance and navigation. Therefore, it is commonly assumed that each vehicle will be equipped with a GPS receiver to obtain location information required for routing purposes and safety application scenarios.

2.3 Dedicated Short Range Communication (DSRC)

Dedicated Short Range Communications (DSRC) is a suite of standards mainly used in VANET safety communication. The fast exchange of safety messages, combined with knowledge about other moving vehicles invisible to drivers extend the safety concepts of VANET considerably [91]. DSRC is a two-way short- to- medium-range wireless communications capability that supports critical data transmission required for cooperative active safety applications [12]. The U.S. Federal Communications Commission (FCC) allocated 75 MHz of a freely licensed spectrum in the 5.9 GHz band for use by ITS vehicle safety and mobility applications. However, the European Telecommunications Standards Institute (ETSI) allocated 30 MHz in the same band. The U.S. DOT commits to the use of the DSRC technologies for both V2V and V2I active safety applications because DSRC is the only available technology that fulfills the latency, accuracy, and reliability requirements of these applications [12]. DSRC is preferred over WiFi because the huge expansion in the usage of WiFi devices and hot spots could cause uncontrollable levels of interference which could hinder the reliability and effectiveness of safety applications. Also, the typical use cases of WiFi are sparse deployment with stationary channels. However, vehicular communication is required among vehicles, even those moving at a high speed, with multipath fading channel, and often in dense environments [145]. Thus, the DSRC is based on an "association-less" version of IEEE 802.11a standard identified as IEEE 802.11p. The IEEE 802.11 standard is chosen as a basis in order to benefit from its ad-hoc mode. This ad-hoc mode resembles vehicle-to-vehicle communications and hence, simplifies the development of DSRC [91]. In addition, the wide availability of IEEE 802.11a chipsets will facilitate producing DSRC enabled devices [145].

DSRC can provide a data rate of up to 27 Mbps within 1 km by using a two way line-of-sight short-range radio. The cost of DSRC is lower than that of cellular, WiMax or satellite communications [91]. However, DSRC is not expected to replace other wireless technologies nor support all vehicular communication needs. DSRC is envisioned as the main communication technology for safety, short-range applications, subscription free services, road toll services, and other similar localized applications [75]. In fact, a strong research trend in vehicular networks is moving toward utilizing multiple different technologies to create *heterogeneous vehicular networks* [154]. The motivation behind this trend is that each technology offers unique benefits. WiFi, for example, would encourage the integration of other road users such as cyclists and pedestrians into the vehicular network. Cellular technology is widely available and designed for delivering large amounts of data over wide coverage. However, there is no consensus concerning how to interface different technologies with the applications [41].

The higher layers of the protocol stack are defined in a suite of standards known as IEEE 1609 Wireless Access in Vehicular Environments (WAVE). This suite addresses security (IEEE P1609.2), networking and messaging (IEEE P1609.3), and channel management (IEEE P1609.4). In particular, IEEE P1609.3 defines a WAVE Short Message Protocol (WSMP) that allows a vehicle to beacon messages in the local vicinity. WSMP also allows carrying messages on both the control and service channels. The applications can directly control the lower-layer parameters such as transmit power, data rate, channel number and receiver MAC addresses through WSMP [84]. Furthermore, the WSMP packet is significantly reduced with an overhead of 5-20 bytes, compared to a minimum of 52 bytes of a UDP/IPv6 packet. As shown in Figure 2.3, WAVE architecture uses IEEE 802.11p for physical and MAC layers in addition to IEEE 1609.4 to support the multichannel operations in the MAC layer. The WAVE architecture supports two protocol stacks to accommodate both stringent communications through WSMP and traditional data exchanges through TCP and UDP protocols over IPv6. Both stacks use the same physical and data link layers but differ from each other in the network and transport layers. Additionally, WAVE has a management entity in a management plane corresponding to each layer in a data plane which is used in system configuration and maintenance [77].

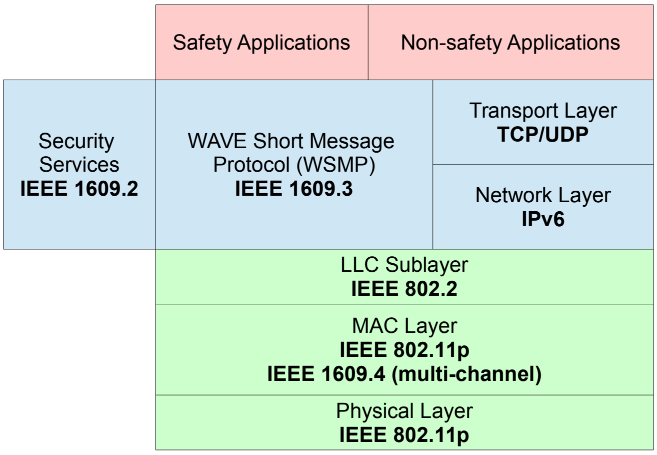

Safety Applications
Non-safety Applications
Security Services IEEE 1609.2
WAVE Short Message Protocol (WSMP) IEEE 1609.3
Transport Layer TCP/UDP
Network Layer IPv6
LLC Sublayer IEEE 802.2
MAC Layer IEEE 802.11p IEEE 1609.4 (multi-channel)
Physical Layer IEEE 802.11p

Figure 2.3: WAVE protocol stack

2.4 Security and Privacy in VANET

Security and privacy issues become more challenging in VANET due to the unique characteristics of VANET. On the one hand, malicious behaviors, such as injecting false information into the network, could be dangerous to users [137]. For example, if a vehicle falsely reports a sudden accident on the road, drivers of nearby vehicles may react incorrectly as they cannot actually see the accident. This situation may in turn cause a real accident. In non-safety applications, an attacker may report false traffic jams on his road and make vehicles take other roads which would lead to low traffic volume on his road [110]. In addition to providing the protection against different types of attacks, trusted traffic authorities should be able to trace and reveal the identity of message senders as an aid in identifying reasons for accident or finding accident witnesses. Therefore, user authentication, authorization and data trust must be included in VANET. Moreover, latency constraints of VANET applications pose more challenges in case cryptographic techniques would be used. These techniques must not increase the communication and processing overhead. On the other hand, user privacy is a crucial issue in VANET. The sensitive or identifying information, such as license plate number, vehicle position and traveling routes, must be well-protected. To ensure privacy, user anonymity should be maintained when vehicles provide information to the network. Otherwise, attackers eavesdropping on the wireless medium can track a vehicle and link its movements with the actual identity. The consequences of this tracking scenario could be annoying targeted ads, movements surveillance and disclosure of sensitive places. The contradiction between anonymity and identity traceability forms an essential challenge for security and privacy.

Although of theses challenges, some VANET characteristics support security and privacy techniques [103]. Vehicles are subject to regular inspections which facilitate update of existing software, download of new certificates and scanning of the system to identify viruses and worms. Moreover, all vehicles must register in a central authority by default. Vehicle registration makes assigning keys to vehicles much easier and more secure. Furthermore, law enforcement mechanisms which support securing the networks against detected attacks already exist in the transportation system.

2.4.1 Requirements

Security and privacy techniques in VANET need to satisfy a set of requirements. Schaub *et al.* [116] categorized these requirements into basic, security and privacy requirements. Basic requirements are those arising from the unique characteristics of VANET. Security requirements are those required to protect the network and its entities from possible attacks and misuses. Privacy requirements are those required to protect the identity of drivers and their movements from potential misuses by unauthorized entities. Several research works, surveys and standards studied and discussed different security and privacy requirements [9, 48, 58, 62, 109, 111, 137]. We merged and organized these requirements in light of the analysis conducted by Schaub *et al.* [116] as follows:

Basic Requirements

- **Real time constraints.** Due to the high mobility and frequent topology changes, the communication window between vehicles is very short. Also, safety applications must respond quickly to the received warning messages. Therefore, it is crucial to minimize the communication and processing overhead.

- **Robustness and availability.** VANET must be robust and provides its services despite the expected high mobility, frequent topology changes and security attacks.

- **Scalability.** On long term basis, VANET will compromise millions of vehicles; therefore, applications and mechanisms should be scalable to handle a large number of nodes.

- **Initial sparse environment.** VANET will be gradually deployed [31, 64] and thus mechanisms should work autonomously without depending on the existence of a fully deployed infrastructure. Thus, applications should be able to provide their services in a sparse environment such as low penetration rate among vehicles and sporadic infrastructure access.

- **Support of various communication patterns.** VANET applications use various communication patterns such as broadcasting and geocasting, along with communication with infrastructure which may be unicast or multicast. Security and privacy mechanisms should consider and support these different patterns.

Security Requirements

- **Authentication.** Authentication is generally used to verify the genuineness of certain claims. Authentication in VANET includes both sender authentication to verify the legitimacy, and message integrity to ensure that the message was not modified since it was sent. For privacy purposes, the real identity of the sender should not be exposed during the verification process and thus anonymous credentials should be used instead.

- **Accountability.** Since vehicles are authenticated, they are accountable to legal authorities for messages they send. Accountability implies non-repudiation which means the sender cannot deny having sent the message (non-repudiation of origin) or the recipient cannot deny having received the message (non-repudiation of receipt). This requirement is also applied when anonymous credentials are employed. Trusted authorities should be able to map anonymous credentials (pseudonyms) to their real identities for law enforcement and liability purposes.

- **Restricted credential usage.** When anonymous credentials are used, they have to be restricted and controlled by an authority to prevent impersonation (acting as another user) and Sybil attacks (using several credentials in parallel to act as several users simultaneously). Also, the validity period of a credential must be limited to prevent an adversary from accumulating credentials for Sybil attacks. However, using short validity periods increases the number of required credentials to be loaded into vehicles and the frequency of loading them.

- **Credential revocation.** The misbehaving vehicles must be prevented from using the network through revocation of their credentials. The essential issue of credential revocation is the scalability and efficiency of the revocation method.

- **Authorization.** Roles must be assigned to vehicles based on type and capability to define what is allowed and what protocols to execute [102]. For example, private vehicles have to be prohibited from sending emergency messages.

- **Confidentiality.** Safety and traffic-related messages should not be encrypted so that they are available to everyone [9]. However, in some cases, such as group communication and key exchange process, the data should be encrypted to prevent unauthorized access. In general, when entity identification is required, communication should be kept confidential [17].

- **Data trust.** Even with authenticated honest users, malfunctioning sensors, invalid aggregation or malicious applications can provide inaccurate information to other vehicles leading to wrong decisions. Thus, disseminated information should be evaluated against its accuracy and trustworthiness. This requirement cannot be achieved by traditional cryptographic techniques but rather by measuring the reported events and providing a credibility or plausibility rank in real-time. Therefore, vehicles should be able to discard messages from revoked or untrusted nodes.

- **Attack prevention rather than detection and recovery.** Security should focus on preventing attacks rather than detecting them and alarming users to take actions. For example, in safety applications, attack detection and warning arrive too late for the user to take an appropriate action [103].

Privacy Requirements

Privacy requirements are usually considered for private vehicles rather than for RSUs and public vehicles such as emergency vehicles and buses.

- **Minimum disclosure.** Information disseminated from vehicles during communication should be kept to the minimum. Information disclosure has to be adaptive to application requirements, as coarse as possible and as detailed as necessary.

- **Conditional anonymity.** Anonymity means not only the sender identity should be kept unknown, but also a message cannot be linked to a specific vehicle using its content. However, vehicles are not totally anonymous in nature as their license plate numbers are still visible. Legal authorities should be able to reveal the identity of the sender of a message when needed.

- **Unlinkability.** Unlinkability means two or more items of interest cannot be linked together. Items of interest can be, for example, messages, credentials or vehicles. Depending on the item of interest, unlinkability can refer to other privacy concepts. For example, unlinkability of a sender to a message it sent is equivalent to sender anonymity. Unlinkability of a message to its originator is equivalent to untraceability. Unlinkability of consecutive messages from one vehicle is equivalent to tracking immunity.

- **Distributed resolution authority.** It is desirable to distribute the ability to reveal user identity to several trusted authorities. This ensures that no single authority can misuse the resolution ability in case of hijack or corruption. Distributed resolution authority makes it more difficult to launch attacks targeted at trusted authorities.

- **Perfect forward privacy.** Revealing the identity of a specific credential should not lead to or help in revealing further credentials of the same user.

Interrelations of requirements

The discussed requirements pose interrelations and conflicts among each other. One of the important design issues in designing a VANET security and privacy technique is to handle these conflicts and trade-offs [116]. Since the basic requirements are obtained from VANET characteristics, they effect all other requirements indirectly. Security and privacy requirements place constraints on each other.

Authentication and accountability are limited by anonymity; the user identity must be unknown but authenticated by other users during communication. User anonymity must be preserved from possible abuse by authorities; hence, resolution ability should be distributed over several entities. In the same way, accountability must guarantee unlinkability among users' pseudonyms and perfect forward privacy. Restricted credential use and revocation of credentials are derived from and strengthened by authentication and accountability. Minimum disclosure, anonymity and unlinkability requirements support each other. Distributed resolution authority and perfect forward privacy do not prevent accountability but also do not allow more information to be revealed than required for resolution and support minimum disclosure. In conclusion, meeting all these constrained requirements at once is challenging for any security and privacy technique [116].

2.4.2 Attack and Threat Models

In this section, security attacks threatening VANET are discussed. By design, VANET inherits all known and unknown vulnerabilities of MANET; security issues in VANET, however, are more challenging due to its unique characteristics and contradicting requirements. Raya and Hubaux [111] classified the capacity of an attack into four dimensions:

1. **Insider vs. Outsider.** The insider attacker is an authenticated user in the network who owns a certified key and can communicate with other members. The outsider is considered as an intruder and has less privileges than the insider which in turn leads to less threats.

2. **Malicious vs. Rational.** A malicious attacker aims to harm other members or the functionality of the network. A rational attacker seeks personal benefits and hence her means and target are more predictable.

3. **Active vs. Passive.** An active attacker may inject packets or signals into the network, modify relayed messages or jam communication. On the contrary, the passive attacker eavesdrops on the wireless medium to learn information about the system entities without affecting them.

4. **Local vs. Global.** When an attacker is limited in scope even if she compromises several vehicles or RSUs, it is called a local attack. An extended attacker can control several entities scattered across the network.

Threat models subject to vehicular networks are extensively studied in literature [55, 102, 103, 109, 153]. Next, we briefly list potential threats.

- **Bogus information.** An attacker diffuses incorrect information to affect the behavior of other drivers. For example, a driver may try to broadcast emergency vehicle warnings to free her road. This is usually a rational active insider attack.

- **Sensor data faking**. An attacker tries to alter the data perceived by a local sensor such as location, speed and direction, to escape liability. This is a local rational active insider attack. In this case, the use of Tamper Proof Device (TPD) that handles attaching such data to messages far from applications is suggested.

- **Denial of service (DoS).** An attacker floods or jams the wireless channel with artificially generated or dummy messages in order to bring down the network. VANET is more vulnerable to DoS attacks due to real-time constraints of its applications [103]. An initial mitigation for this attack is to switch to another wireless channel or even to another wireless technology (celluar or Bluetooth) [111].

- **Movement tracking.** An attacker who eavesdrops on a wireless channel over large parts of the network for a range of purposes such as disclosure of vehicle identity and finding their places of interest. Generally, this attack is malicious passive and global.

- **Message replay attack.** An attacker re-injects previously received beacons to poison a vehicle's location table [153].

- **Message modification attack.** An adversary tries to change the source or the content of a message during or after transmission aiming to escape the liability.

- **Message suppression attack.** An adversary may use one or more vehicles to selectively drop packets from the network such as congestion alerts to make other vehicles enter congested traffic. Similarly, the attacker can drop all received messages forming a sinkhole in the network.

- **Masquerading.** An attacker maliciously or rationally tries to pretend it is another authenticated vehicle by using a false identity.

- **Sybil attack.** An attacker uses a large number of pseudonyms at the same time to pretend it is actually hundreds of vehicles in order to persuade other vehicles there is a traffic jam ahead they should take an alternative route.

- **RSU replication attack.** An RSU can be compromised so that it can be relocated to make other attacks, such as broadcasting false information, in its new location.

- **GPS spoofing.** Since all vehicles include a GPS receiver, an attacker can act as a GPS satellite simulator to generate signals that are stronger than genuine ones [153]. Thus it fools other vehicles by producing false location readings which means most location-based services will work incorrectly.

2.5 Security Approaches

Numerous research works published in the past decade that address security and privacy issues in VANET and have been recently reviewed by Petit *et al.* [105] and Qu *et al.* [109]. The large diversity of the proposed mechanisms results from the trade-offs between security, privacy, efficiency and trust. However, there is a consensus towards adopting public key infrastructure (PKI) for securing VANET [78]. This security approach was initially proposed by Papadimitratos *et al.* [100] during the SeVeCom project and adopted by standardization bodies (ETSI TS 102 941 [10] and IEEE 1609.2 WG [11]). However, there are other VANET security architectures that are based on different approaches such as identity-based cryptography, group signature and symmetric cryptography, as discussed next.

2.5.1 Public Key Infrastructure

In the conventional PKI, each node has private and public keys to authenticate messages. A certification authority (CA) is required to certify public keys and announce revoked nodes. Although the PKI fulfills many VANET security requirements, it should be modified to support the privacy requirements. For example, certificates should not contain any identifying information about the owner. Also, keys should be changed periodically to avoid linking the signed messages by the same certificate. Therefore, Raya and Hubaux [110] proposed that each vehicle should be provided with two types of certificates: 1) unique long-term identity and a key pair and 2) several pseudonyms associated with anonymous key pairs. The long-term certificate is issued by a CA and should be installed securely into the vehicle. The anonymous keys are also certified by the CA and are used in signing messages. To allow message verification, pseudonym certificates must be sent along with messages. Thus, receivers can authenticate messages without revealing the identity of the sender. A vehicle uses a pseudonym for a period of time then switches to another, not previously used pseudonym. A tamper proof device (TPD) is embedded in vehicles to generate key pairs and send public keys to the corresponding CA for certification [100]. The CA signs the public keys, generates pseudonyms and stores them with the vehicle's long-term identity. Each pseudonym certificate contains an identifier of the CA, the lifetime of the pseudonym, the public key, and the signature of the CA. The TPD manages received pseudonyms and ensures that only one pseudonym is used at a time to prevent Sybil attacks. The CA revokes pseudonyms of the misbehaving vehicles by broadcasting a certificate revocation list (CRL) to the network.

There are many challenges that appear with this security approach. First, who will authenticate vehicles in bootstrapping: the transport registration authority or the manufacturer. Second, how are the certification authorities organized and how is the mapping of keys to the real identities maintained for accountability and liability purposes. Should it be centralized across regions in a central CA or hierarchical based on the regional structure. Both options pose different challenges. The central management challenges the scalability while the hierarchical management poses questions about the recognition of certificates issued by different authorities. Third, how and how many pseudonyms should be loaded into vehicles. Will they be requested by vehicles online or downloaded periodically during vehicle check ups. When should pseudonyms be changed: at random periods, in mix zones, in social spots or preceded by a silent period. Fourth, how are revoked pseudonyms published: through RSU, forced by a TPD or using revocation lists. Fifth, how is resolution authority technically guaranteed to be distributed on multiple entities. Last but not least, does this architecture support the stringent latency constraints of safety applications. This security architecture is studied intensively in literature such as [14, 78, 100, 64].

2.5.2 Identity-based Cryptography

Identity-based cryptography (IBC) [24] is a type of asymmetric cryptography in which any vehicle can form the public key from its corresponding identity string. The main benefit of IBC is the elimination of the need to certify the public key and exchange certificates within messages. However, a centralized trusted authority, which owns a master private key, is needed to generate a private key for each vehicle. Thus, the vehicle legitimacy is implicitly guaranteed, rather than explicitly verified by a certificate, because only an authorized vehicle would receive a private key corresponding to its identity. The IBC communication and storage overheads are significantly reduced compared with the PKI-based approach. Instead of using the vehicle identity, the trusted authority generates and sends pseudonyms to each vehicle along with their corresponding private keys. Since any vehicle can generate the public key of a pseudonym, no additional information is required to be attached to the message. The main drawback of IBC schemes is the reliance on a centralized trusted authority for private key generation. VANET security mechanisms based on IBC can be found in [15, 22, 126].

2.5.3 Group Signature

In group signature [34], each member in a group has a private key to sign a message anonymously on behalf of the group. Other members use the shared group key to verify signed messages without revealing who signed them. However, a group manager can use its key to reveal the original signer of a message. Additionally, two messages signed by the same vehicle cannot be linked together because group members cannot determine if those messages came from the same or different members. By design, group signature supports anonymity, untraceability, unlinkability and unforgeablity (non-members cannot produce authenticated messages). Therefore, there is no need for generation, storage, verification and revocation of numerous pseudonym certificates per vehicle as in PKI and IBC approaches. Despite these appealing features, there are several challenges. Similar to IBC, the verification and authentication processes are time consuming. Second, group formation, members revocation and inter-group communication are essential issues for a successful realization in VANET. Security mechanisms based on group signature can be found in [63, 86, 127]. Hybrid mechanisms that utilize group signature partially are discussed in [30, 88].

2.5.4 Symmetric Cryptography

In symmetric schemes, a Message Authentication Code (MAC) is used for message authentication. The sender hashes the message and a secret key. Any receiver must know the secret key to verify the MAC by performing the same operation on the message. Thus, any node with knowledge of the secret key can generate valid MACs, but the sender accountability is not provided. The main benefits of this approach are the fast encryption and decryption times as well as less security overhead. In addition, the key distribution mechanism could be simpler and cost less than the deployment and maintenance of a PKI scheme. However, a reliable symmetric scheme requires that exposure of single or some secret keys should not compromise authentication of all vehicles. Xi *et al.* [146] proposed the symmetric random key set approach. In this scheme, sets of symmetric keys are drawn from a shared key pool and one key is shared by several vehicles. Thus, the identity and the keys are not closely correlated. This is helpful in key revocation because even if some of the keys have been revoked, the rest of the vehicles can still be authenticated. Hu and Laberteaux [70] applied the TESLA symmetric authentication protocol, which does not require RSU support as in [146]. In TESLA [104], signers use symmetric keys derived from hash chains for message authentication and release keys after a certain period of time. A message is authenticated with a key that has not been released yet, thus, receivers must store messages until the corresponding key or a higher key has been released. In [70], key release periods are determined according to the message frequency and the allowed latency.

2.6 Privacy Approaches

Privacy is preserved by achieving anonymous communication which should be sufficiently robust against different de-anonymization attempts whether from internal or external entities. Anonymity is often a method to protect privacy, as well as a goal in itself [17]. Satisfying privacy requirements, discussed in Section 2.4.1, depends on the employed security approach. For example, if the group signature mechanism will be used in message authentication, anonymity and unlinkability are implicitly guaranteed within the group. However, if the PKI mechanism will be employed, additional privacy mechanisms are required such as using pseudonyms for the anonymity requirement and changing it periodically for the unlinkability requirement.

Since there is a growing consensus towards adopting PKI for securing VANET [78], we focus on privacy mechanisms for this security approach. As discussed in Section 2.5, pseudonyms are used instead of long-term certificates to provide anonymity. Pseudonyms were originally introduced by Chaum for anonymity of electronic transactions and defined as "a public key used to verify signatures made by the anonymous holder of the corresponding private key" [35]. Pfitzmann and Hansen defined a digital pseudonym as "a bit string which is unique as identifier (at least with very high probability) and suitable to be used to authenticate the holder's item..." [106]. Since a pseudonym is unique, all its authenticated messages are linkable. To provide unlinkability, a vehicle uses a set of pseudonyms such that a pseudonym is used for a short period of time. Based on these definitions and features, Petit *et al.* [105] identified pseudonyms requirements in order to ensure privacy requirements as follows:

- **Uniqueness**. It is guaranteed by the pseudonym provider and the underlying security mechanism used to generate the pseudonym.

- **Availability** A new pseudonym should always be available for the vehicle in case of pseudonym change. A new pseudonym can be provided by storing a large set of pseudonyms in the OBU or through a dynamic pseudonym refilling mechanism.

- **Time-limited**. A pseudonym must have a validity period to avoid tracking messages. This time limit is ensured by the signed certificate that accompanies the pseudonym.

- **Pseudonym change block**. The ability to prevent pseudonym change is sometimes required to ensure resilience against depletion attacks and preserving safety level.
- **Link to other identifiers** When a pseudonym is changed, all the other identifiers (IP and MAC addresses) used by the same vehicle have to be changed as well.

To prevent linkability of messages, a vehicle must change pseudonyms; thus, an adversary could only link a few messages. However, pseudonyms should be changed in appropriate contexts to avoid trivial linkability between old and new pseudonyms. For example, if a vehicle changes its pseudonym alone in a small area, the adversary can guess an event of pseudonym change and relink them. Simultaneous pseudonym changes are not necessarily sufficient, unless the trajectories of vehicles are unpredictable by the adversary, as will be shown in Chapter 3. Numerous research works consider how, where, and in which situations pseudonyms should be changed in order to be effective. The pseudonym change mechanisms can be categorized into five groups: periodical, context-based, in a mix-zone, after a silence period, and collaborative. When a mechanism employs two or more techniques, we categorize it according to its main contribution. In the rest of the thesis, we refer to pseudonym change mechanisms as *privacy schemes* .

Most of the privacy schemes assume a worst-case adversary who can eavesdrop all exchanged messages, especially safety beacon messages. Since these messages are broadcast frequently and contain spatiotemporal information about vehicles, linking consecutive messages of new and old pseudonyms is effectively attainable using target tracking techniques [45, 143]. The privacy level is measured using different metrics such as the anonymity set size, entropy and the probability of tracking success. The lack of consensus on a standard privacy metric for vehicular networks makes a comparison of different schemes difficult [105]. We will provide detailed evaluation and comparison among several privacy schemes in Section 7.6.

2.6.1 Periodical Change

Periodical schemes change pseudonyms at fixed or random times. Fixed periods may increase simultaneous pseudonym changes among nearby vehicles although an adversary would be able to predict when pseudonyms would be changed. Random periods overcome this prediction issue. Eckhoff *et al.* [42] proposed a time-slotted pseudonym pool with a swapping capability. Each vehicle is equipped with a pseudonyms pool whereby each pseudonym is used for a specific time slot. When all pseudonyms are used, a vehicle starts using the pseudonym of the first time slot. Vehicles can swap their pseudonyms valid for a specific time slot to ensure each vehicle has only one pseudonym in each time slot. Swapping of currently used pseudonyms is done by carefully investigating the context information, such as speed, heading, positions of other vehicles. Swapping is performed only if the environment information leads to improving the anonymity of both vehicles. This scheme eliminates the mapping between pseudonyms and real identities which disables the accountability requirement. Freudiger *et al.* [52] proposed initiating a pseudonym change when it is considered old and there are other vehicles in proximity. The age of pseudonym is measured by a linearly increasing function of time and reset to zero after a successful change. The authors calculated the probability distribution of the pseudonym age analytically under the assumption that an adversary becomes confused if two or more vehicles change their pseudonyms followed by a silence period. Freudiger *et al.* also studied the probability of cooperation (i.e., at least one neighbor changes its pseudonym). They found that the probability of cooperation increases logarithmically with the increase of pseudonym aging rate and decreases with the larger vehicles meeting rate. Pan *et al.* [98] presented an analytical model to quantify the expected anonymity set size in random pseudonym change schemes. They analytically computed the probability of the target vehicle to change its pseudonym simultaneously with its neighbor. Then, they calculated the expected size of the anonymity set. They considered the anonymity set to be the nearby vehicles with similar direction and speed. Thus, each vehicle in the anonymity set is equally likely to be the target vehicle and thus the tracker cannot identify the target. According to the experiment results, the expected size of the anonymity set ranges from 1.04 to 1.12 depending on the pseudonym change period. These results are obtained when there are four neighbor vehicles around the target during the time it changes its pseudonym. This result is important because it shows that changing pseudonyms randomly does not provide enough anonymity even if the traffic is dense.

2.6.2 Context-based

In the context-based approach, a vehicle changes its pseudonym based on context parameters whether internal parameters such as the current speed and direction or external parameters such as the density of the surrounding traffic. Raya and Hubaux [110] proposed changing pseudonyms when the adversary cannot correlate the old and new pseudonyms. They calculated a lower bound for pseudonym lifetime based on the vehicle transmission range and the distance over which a vehicle does not change its speed and lane. Based on this lower bound, they estimated that approximately 43800 keys are required per year to be loaded in a vehicle (assuming 2 driving hours per day). Li *et al.* [83] proposed two protocols: Swing and Swap. In Swing, vehicles change pseudonyms when changing their velocity (speed and direction). To increase the probability of simultaneous changes, a vehicle first checks that there is at least one vehicle in its vicinity and broadcasts its intention to change its pseudonym to the nearby vehicles. In Swap, vehicles exchange their pseudonyms with probability 0.5 before a random silent period. Swapping pseudonyms increases the anonymity set by including vehicles that have not changed their pseudonyms with the vehicles doing the pseudonym change; the adversary does not know which vehicle(s) exchanged pseudonyms or if the vehicle exchanged its pseudonym at all. The Swap scheme challenges the accountability requirement since the mapping between the pseudonyms and the real identity is not updated in the central authority.

Gerlach and Guttler [57] proposed the concept of *context mix* where a vehicle changes its pseudonym if there are N neighbors within a small radius (4.25 m) after holding the last pseudonym for a specific stable time (1 min). The vehicle assesses the situation after each change to ensure it is successful, that is, other vehicles changed their pseudonyms as well. If this is not the case, the vehicle restarts the change cycle. Gerlach and Guttler employed a tracker that fails if two or more similar vehicles changed pseudonym simultaneously. Based on their experiments, fewer vehicles are tracked when they change pseudonyms in mix contexts than if they change at random periods. Buttyán *et al.* [28] proposed the SLOW protocol which stops sending messages when the vehicle's speed drops lower than a preset threshold. If a vehicle remained silent for a while, it changes its pseudonym. The idea behind choosing low speed is that it is less likely to cause fatal accidents and indicates a natural mix areas where many vehicles are located in close proximity. Buttyán *et al.* assumed a global observer which tracks vehicles by predicating the next position based on information included in the last two beacons. The observer has knowledge of probability distribution of traffic flow and time delay through road intersections. The privacy level is measured by finding the percentage of vehicles that are tracked completely in the simulated traces. Based on their results, the tracking effectiveness is reduced when vehicles pass through several intersections with silent periods; it depends on the speed threshold and vehicle density.

Lu *et al.* [87] proposed to change pseudonyms in social spots such as signaled intersections and parking areas where several vehicles are stopped for a period of time. Before leaving a social spot, vehicles change their pseudonyms to create a dynamic mix zone. Lu *et al.* proposed a self-delegation key generation model where the driver can generate short-life pseudonyms using an authorized anonymous key provided by a trusted authority. This model allows vehicles to flexibly change their pseudonyms frequently. The experiment

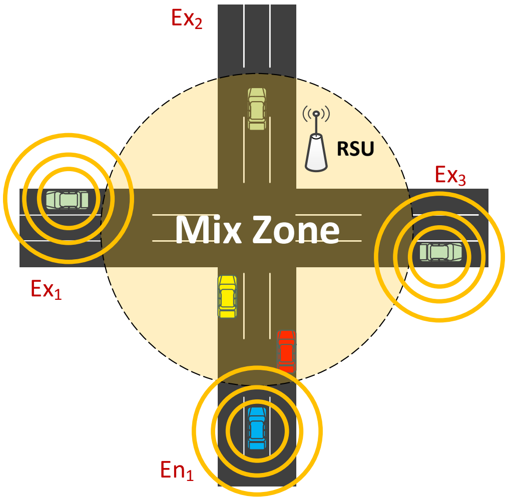


Figure 2.4: A mix zone at an intersection controlled by an RSU. The adversary cannot observe messages broadcast within the mix zone.

results show an increase in the anonymity set size with the increase of the vehicle arrival rate and the stopped time at social spots. However, this scheme ignores the position precision of safety messages. Such precise spatial information can distinguish vehicles stopped at social area. Also, not all social spots are a perfect place for changing pseudonyms. Social spots such as shopping malls are considered places of interest of the driver and the adversary considers them the end of the trip. Thus, the adversary may not be interested in linking pseudonyms before and after such spots.

2.6.3 Mix Zone

A mix zone was first introduced by Beresford and Stajano [18] for preserving location privacy. This approach is analogous to a mix node of a mix network [35], which changes the order of messages and their encoding to make linking the message sender and receiver difficult. In VANET, the mix zone makes it difficult for the adversary to link the vehicles that exit from the mix zone to those that entered it earlier. Figure 2.4 illustrates a mix zone controlled by an RSU at a road intersection. In a mix zone, the adversary cannot observe broadcast messages and thus cannot predict the movement of the vehicles. If vehicles would change their pseudonyms within the mix zone, the adversary cannot correlate leaving vehicles with those entering the zone earlier. Hiding messages in a mix zone is realized by keeping silence [27] or by encrypting messages using a shared key obtained from an RSU [53].

Buttyán *et al.* [27] introduced the concept of mix zone in vehicular networks. The authors assumed an adversary who knows the conditional probability of leaving the mix zone at port j given that the entry point was port i. The adversary calculates the probability distribution of the time delay when traversing the mix zone between each pair of ports. These probability distributions are obtained by monitoring vehicle traffic at intersections. To correlate leaving and entering vehicles, the adversary monitors leaving vehicles and assign them to entering vehicles where the correlation probability is maximum. Buttyán *et al.* showed by simulation that the tracking success increases with the increasing number of attacker receivers at intersections. However, there is a saturation point when the adversary covers only half of the intersections. These authors also observed that the success probability of the tracker is nearly independent from the traffic density above a given tracker strength. Freudiger *et al.* [53] realized mix zones using symmetric cryptography and introduced cryptographic mix zones (CMIX). The basic idea of CMIX is that vehicles obtain a symmetric key from the RSU of the mix zone and encrypt all messages while passing by the zone. Keys are also forwarded upon request from vehicles outside the range of RSU to be able to decrypt received messages from vehicles within the zone. Ying *et al.* [149] proposed a scheme called dynamic mix zone for location privacy (DMLP). In this scheme, a mix zone is dynamically formed at the time the vehicle requests it with the aid of RSUs and control servers. DMLP encrypts all transmitted messages while the vehicle is within the mix zone. The size of the mix zone is determined by the vehicle's predicted location, the traffic statistics and the level of vehicle's privacy requirement.

Choosing the effective places to deploy mix zones is a challenging problem which has gained large consideration in literature. Freudiger *et al.* [54] proposed an algorithm to find the optimal placement of mix zones by maximizing the mixing effectiveness of the system at an affordable cost for mobile nodes. The algorithm ensures a lower bound location privacy by enforcing a maximum distance between traversed mix zones. Freudiger *et al.* also proposed a new metric based on the mobility profiles. In this metric, the traffic at an exit point is modeled as the conjunction of the flows initialed from all entry points, then the probability of error of the adversary in assigning an exiting node to the correct flow is computed. Similarly, Sun *et al.* [128, 129] proposed a statistics-based metric for evaluating the effectiveness of a mix zone. This metric is employed to determine the fewest mix zones that guarantee vehicles at any place pass through a mix zone in a certain driving time and a small extra overhead of adjusting routes. Palanisamy *et al.* [94, 95, 96] proposed the MobiMix framework which is a construction and placement model for mix zones that is robust against timing and transition attacks. This model takes into account multiple factors in constructing and placing mix zones, such as the road topology characteristics, and the timing and the transitioning probability of vehicles in terms of their movement trajectory. The authors also provided a formal analysis on the vulnerabilities of directly applying the rectangle mix-zones to road networks in terms of anonymization effectiveness and resilience to timing and transition attacks.

2.6.4 Silent Period

The silent period approach can be considered as a special type of mix zone where it is not necessary to place the zone in fixed locations. Huang *et al.* [71] proposed entering a silent period before a pseudonym change to harden tracking, especially in highly dense spots such as intersections or traffic lights. However, silent periods conflict with delay-sensitive safety applications which are mostly required in these dense spots. Sampigethaya *et al.* [114, 115] applied silent period in VANET when vehicles are merging and/or changing lanes when joining or leaving a freeway. The ramps that allow vehicles to merge into lanes on freeways are relatively safer locations compared to the main lanes of freeway [114]. These authors also proposed group communication with silent period for V2I communication. Each vehicle group has a group leader who acts as a proxy to all the group members. The group leader can broadcast aggregated traffic information of the group while the other members are silent. Burmester *et al.* [26] showed vehicle tracking before and after silence periods using Bayesian analysis. They claimed that the complexity of the road topology, the traffic density, the vehicle proximity and the unpredictable behavior of drivers are the main factors to harden linkability. For this purpose, they concluded that pseudonyms should only be updated when a vehicle crosses a joint point during which a short period of silence takes place.

2.6.5 Collaborative

In the collaborative approach, nearby vehicles communicate with each other to synchronize their pseudonyms change to increase adversary confusion. Liao and Li [85] extended the context mix approach proposed in [57] to have synchronous pseudonym change with two or more similar vehicles. They proposed to set a flag included in beacons when the minimum stable time of the pseudonym expires. The vehicle then waits until receiving beacons of k vehicles that have similar status and a set flag as well. The experiment results show that the synchronous pseudonym change increases the number of successful changes and reduces the number of pseudonyms used by vehicles to a greater extent than the mix context approach. Wasef and Shen [138] proposed a random encryption period (REP) scheme which employs encryption to form a secure group among vehicles to change their pseudonyms. A vehicle intending to change its pseudonym communicates with nearby vehicles and arranges a period of time in which all messages are encrypted and pseudonyms are changed. However, an active attacker may participate in the encryption period and can therefore observe the pseudonym change [117]. Pan and Li [97] proposed a cooperative pseudonym change scheme based on the number of neighbors. Vehicles monitor their neighbors within radius R and wait until they reach a threshold k. When this trigger occurs, the vehicle sets an internal flag, broadcasts it within its beacon and changes its pseudonym in the next beacon. When a vehicle receives a beacon with a set flag or its internal flag is set already, it changes pseudonym immediately. The results of the experiment show that the expected size of the anonymity set increases with the increase of traffic density and the radius R; it decreases with the increase of threshold k. The anonymity set is expected to increase with multi-lane roads. Pan and Li compared this scheme with a non-cooperative scheme which changes pseudonym once k neighbors were detected. They showed that the enhancement of cooperative scheme over the non-cooperative scheme increases until the average number of neighbors of the target vehicle approaches the threshold k and then it decreases.

2.7 Summary

In this chapter, the underlying theories behind this dissertation are discussed. First, the vehicular network is introduced explaining its applications, characteristics and enabling wireless technology and protocols. Then, the security and privacy requirements and possible threat models are presented. Last but not least, different security and privacy approaches are categorized and surveyed.

3 Multi-Target Vehicle Tracker

3.1 Introduction

As indicated in Section 1.4, we use an empirical tracker as the adversary model for privacy schemes evaluation. This tracker tries to link subsequent beacon messages broadcast from each vehicle even if these messages are identified by different pseudonyms. Thus, this tracker must be very robust to truly reflect the effectiveness of privacy schemes. Beside the adversary model, vehicle tracking will be used in different aspects of this thesis. We propose using a local tracker inside vehicles to keep track of the movement of nearby vehicles. This local vehicle tracker will enhance the quality of service of safety applications, as will be discussed in Chapter 5. In addition, it can be used to help vehicles improve its location privacy by determining the appropriate context in which a vehicle should change its pseudonym, as will be explained in Chapter 7. Therefore, we discuss vehicle tracking in this chapter to facilitate the discussion in later chapters.

3.1.1 Vehicle Tracker Model

The vehicle tracker collects beacon messages broadcast by vehicles located within the coverage range of its receiver. If the tracker uses multiple receivers distributed over the road network, then it can collect all received messages forming the vehicle traces that passed the covered area. In fact, both tracker models (i.e., with single or multiple receivers) are used in different parts of this thesis. The former model is used in the context monitoring module inside vehicles to enhance their awareness about the surrounding traffic. The latter model is used as an adversary model to measure the privacy level attained by a privacy scheme. Regarding the beacon message, we assume it includes at least a time stamp and the current position, speed and heading of the vehicle. It may also include other vehicle-related measurements and information such as the acceleration and the vehicle type and size. When the tracker collects beacons, it quantizes them according to the default beaconing time t_b. It rounds the time stamp included in the beacons to the nearest beaconing time in order to divide them into time steps. Thus, a new beacon from each vehicle is expected to appear once in every time step. In this chapter and unless stated otherwise, we assume that beacon messages are completely anonymous or, in other words, that a new pseudonym is used with each beacon message. If a tracker can achieve high accuracy in this worst case, it will track vehicles more accurately when a pseudonym is used for several beacons.

3.1.2 Multiple Target Tracking (MTT)

Vehicle tracking using beacon messages can be considered as a typical well-studied multiple target tracking (MTT) problem. The MTT involves comprehensive approaches and algorithms that are employed in several applications [23, 147]. It assumes a set of measurements or observations detected by a sensor in each time period; this set is referred to as a *scan*. Its goal is to find the best estimate of the target states in each scan. Measurements are assumed to be noisy and include clutter caused by false measurements not originating from real targets.

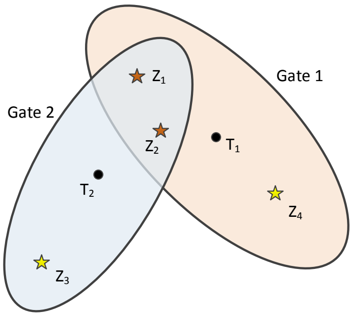


Figure 3.1: Gates of two tracks T_1 and T_2 with three measurements in each. Two measurements Z_1 and Z_2 are located in the intersection of gates. Only measurements located in the tracker gate are considered in the data association process of that track.

The MTT can be explained by gradually investigating tracking cases from simple to complex. The simplest case is tracking a single target with no clutter. The sensor acquires a noisy measurement every time step and it is required to obtain an enhanced target state. Thus, a *state estimation* filter, such as Kalman filter, is employed to combine the acquired measurement and the calculated state obtained from a predefined kinematic model for that target. The estimation filter converges overtime to form a more accurate track for the target than that detected by the sensor. When clutter is present, several measurements are detected in every scan but only one of them is originating from the target, if any. In this case, the estimation filter cannot be used directly as it is unknown

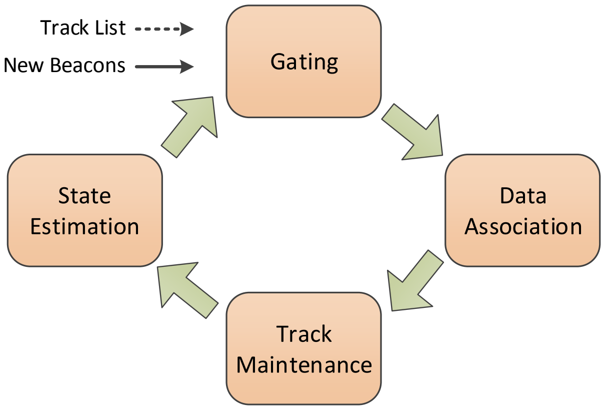


Figure 3.2: Phases of multi-target vehicle tracking.

which measurement belongs to the target. A *data association* process is performed to identify which measurement is most likely originating from the target. However, a validation process or *gating* is performed beforehand to avoid unnecessary computations. Gating forms a validation area around the track and excludes any measurement located outside this area from being tested in the computationally intensive data association process, as shown in Figure 3.1.

The complex tracking case is the multiple target tracking in clutter. Assuming that there is a set of tracks already established for the targets, then, a gate can be formulated around each track. Because these gates can overlap and measurements can be located in more than one gate, as demonstrated in Figure 3.1, the data association process for all tracks must be handled simultaneously. Otherwise, the data association will not be globally optimized leading to false assignments. If the number of targets are unknown or dynamic, a separate or joint process with data association should handle the track initiation, confirmation and deletion, which is referred to as *track maintenance*. Figure 3.2 shows the main phases of MTT that will be discussed in more detail in Section 3.2.

3.1.3 Vehicle Tracking as an MTT Problem

According to the description given in the previous two sections, vehicle tracking in VANET is a typical MTT problem, but it has different assumptions and constraints. First, there is no clutter or false measurements assumed in beacon messages by default. All received messages reflect real vehicles unless the adopted privacy scheme uses dummy traffic. Second, some of detection problems that may occur because of the limitation or deficiency of sensors are unlikely to occur in VANET domain. Examples of these problems are the unresolved measurements problem, which occurs when a single measurement is formed from multiple targets, and the multiple detection problem, which occurs when the same target is detected more than once in a single scan. These problems are considered to be the main challenges for data association [147]. Third, the expected accuracy of beacon information and its broadcast rate are higher than those expected in MTT algorithms. This can be induced by the requirements of safety applications which require precise positions with error less than one meter and a high beaconing rate up to 10 Hz [38]. Fourth, the vehicle movement is predictable and constrained by roads and driving rules which leads to simpler vehicle modeling and tracking. These differences propose that vehicle tracking can be accomplished effectively and efficiently using non-complex MTT approaches and can achieve a high accuracy.

The rest of this chapter is organized as follows. Our proposed tracker and its phases are explained in detail in Section 3.2. In Sections 3.3 and 3.4, we explain the evaluation metric and the experiment results of the tracker, respectively.

3.2 Proposed Vehicle Tracker

As briefly presented in Section 3.1.2, multi-target vehicle tracking consists of four iterative phases: state estimation, gating, data association and track maintenance. *State estimation* (e.g., Kalman filter) is used to obtain an accurate state for vehicles using both inaccurate measurements gained from vehicle sensors and the estimated states obtained from a predefined kinematic model. Because several beacons are received from different vehicles in each time step, *data association* phase is performed to associate the measurements with their originating vehicles. However, a validation phase, or *gating*, is performed prior to data association to prevent unnecessary computations for unlikely associations. Because the number of vehicles is unknown and dynamic, a *track maintenance* phase is needed to handle track initiation, confirmation and deletion.

Next, we will briefly discuss the phases of vehicle tracking. Although it begins logically with gating, state estimation will be discussed first because it is crucial to the remaining phases.

3.2.1 State Estimation

A vehicle state expresses on the set of facts about the vehicle, which include its position, velocity and acceleration. It is practically impossible to determine the exact vehicle state because sensors such as GPS receiver, speedometer, etc. have limited precision and are prune to noise. Thus, in order to track vehicles and link their messages, their state should be better estimated using a state estimation filter. The state estimation filter is not an interpolation or extrapolation but it gives a better estimate for a state $\mathbf{x}_k$ at time k taking into account both the previous states $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$,...,$\mathbf{x}_{k-1}$ and the inaccurate measurement $\mathbf{z}_k$ acquired at time k. The most common state estimation filter is the Kalman filter [76]. The Kalman filter is a set of mathematical equations that provides an efficient iterative method to estimate the state of a stochastic process so that the mean square error is minimized. In order to use Kalman filter in estimating the vehicle state, vehicle dynamics should be modeled in accordance with the Kalman filter model. The Kalman filter assumes that the underlying system is linear where the transition between subsequent states is given by a linear equation. Also, it assumes that the process and the measurement noises are Gaussian distributed. We define the vehicle motion as a linear dynamic model with Gaussian-distributed noise as:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w} \tag{3.1}$$

where $\mathbf{x}_k$ is the vehicle state vector at time step k and $\mathbf{A}$ is the transition matrix that advances the state by one time step. The random variable $\mathbf{w}$ represents the process noise with a normal distribution $\mathcal{N}(0, \mathbf{Q})$ where $\mathbf{Q}$ is its covariance matrix. $\mathbf{z}_k$ denotes the measurement at time step k and is defined as:

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v} \tag{3.2}$$

where $\mathbf{H}$ is the model matrix that maps from the state space to the measurement space. The random variable $\mathbf{v}$ is the measurement noise with a normal distribution $\mathcal{N}(0, \mathbf{R})$ where $\mathbf{R}$ is its covariance matrix. $\mathbf{Q}$ and $\mathbf{R}$ do not change over time.

The state vector $\mathbf{x}_k$ consists of the vehicle position p, speed s and acceleration a in 3D Cartesian coordinates. The transition matrix $\mathbf{A}$ is formulated using motion equations forming a 9x9 matrix. However, such large dimension may lead to inefficiency in computations. It is recommended in [23] to decouple the components of each coordinate because they are independent of each other. Thus, the state vector $\mathbf{x}_k(i)$ and the transition matrix $\mathbf{A}(i)$ of coordinates x, y and z are defined as:

$$\mathbf{x}_k(i) = \begin{bmatrix} p_i \\ s_i \\ a_i \end{bmatrix}, \mathbf{A}(i) = \begin{bmatrix} 1 & t_b & t_b^2/2 \\ 0 & 1 & t_b \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

where the subscript i refers to the x, y or z coordinate, and $1/t_b$ is the beaconing rate. The subscript i is subsequently omitted for simplicity but it is worthy to note that any reference to the state vector $\mathbf{x}_k$ means one part of the vector. We assume that the beacon message contains the current position, speed and heading (i.e., cosine of thetas in each direction) according to the specifications of safety applications [38]. Because the use of the heading in the measurement vector produces a non-linear model, the vectored velocity is calculated using the given heading and the speed. Thus, the measurement vector $\mathbf{z}_k$ and the matrix $\mathbf{H}$ of each coordinate x, y and z are defined as follows:

$$\mathbf{z}_k = \begin{bmatrix} p \\ s \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{3.4}$$

For process noise, we assume that $\mathbf{w} = \begin{bmatrix} t_b^2/2 & t_b & 1 \end{bmatrix}^T$. Thus, the covariance matrix $\mathbf{Q}$ can be defined as:

$$\mathbf{Q} = E(\mathbf{w}\mathbf{w}^T)\sigma_{ap}^2 = \begin{bmatrix} t_b^4/4 & t_b^3/3 & t_b^2/2 \\ t_b^3/2 & t_b^2 & t_b \\ t_b^2/2 & t_b & 1 \end{bmatrix} \sigma_{ap}^2 \tag{3.5}$$

where σ_{ap}^2 is the acceleration variance in the process noise. For measurement noise, we assume that the variances in the measurements of position and velocity (σ_p^2 and σ_v^2, respectively) are provided to the Kalman filter as parameters. Thus, the covariance matrix $\mathbf{R}$ is defined as:

$$\mathbf{R} = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} \tag{3.6}$$

Values of these parameters are carefully selected, as discussed in Section 3.4.1. Thus, the vehicle model is formed and can be used in Kalman filter as shown next.

The Kalman filter is an iterative algorithm and switches between prediction and update steps. At time step k, the prediction step calculates a predicted (a priori) state $\hat{\mathbf{x}}_k^-$ using the estimated state $\hat{\mathbf{x}}_{k-1}$ of the previous time step $k-1$. It also calculates a predicted (a priori) error covariance matrix $\mathbf{P}_k^-$ which indicates the accuracy of the predicted estimate, as specified in (3.7). The predicted state $\hat{\mathbf{x}}_k^-$ is also called *a priori* because it does not include the measurement of the current time step yet.

Prediction Step:

$$\begin{aligned} \hat{\mathbf{x}}_k^- &= \mathbf{A}\hat{\mathbf{x}}_{k-1} \\ \mathbf{P}_k^- &= \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \end{aligned} \tag{3.7}$$

where $\mathbf{A}$ and $\mathbf{Q}$ are matrices defined in (3.3) and (3.5), respectively. It is assumed that the measurements of the first scan initiate the tracks and initialize the state vector $\hat{\mathbf{x}}_0$ at $k = 0$. Also, the initial error covariance matrix $\mathbf{P}_0$ is formed to have a parametric error in position while zero error in velocity and acceleration as follows:

$$\mathbf{P}_0 = \begin{bmatrix} p_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.8}$$

where p_0 is a parameter given to the Kalman filter.

The update step calculates the Kalman gain $\mathbf{K}$ to update the predicted estimate by the observed measurement at the current time step. Also, it computes the residual or innovation $\tilde{\mathbf{z}}_k$, which is the difference between the actual measurement and the estimated one, and the innovation covariance matrix $\mathbf{S}$ which indicates the accuracy of the residual. Both the residual $\tilde{\mathbf{z}}_k$ and its covariance matrix $\mathbf{S}$ are used later in the gating phase.

Update Step:

$$\begin{aligned} \mathbf{S} &= \mathbf{H}\mathbf{P}_k^{-}\mathbf{H}^T + \mathbf{R} \\ \mathbf{K} &= \mathbf{P}_k^{-}\mathbf{H}^T\mathbf{S}^{-1} \\ \tilde{\mathbf{z}}_k &= \mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^{-} \\ \hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_k^{-} + \mathbf{K}\tilde{\mathbf{z}}_k \\ \mathbf{P}_k &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_k^{-} \end{aligned} \tag{3.9}$$

where $\mathbf{H}$ and $\mathbf{R}$ are matrices defined in (3.4) and (3.6), respectively, and $\mathbf{I}$ is the identity matrix. More details about Kalman filter and its derivations can be found in [140].

3.2.2 Gating

Assuming a track is established for each vehicle, a measurement-to-track association should be performed to assign the new measurement to the correct track. Prior to the association, a gating process is required to eliminate unlikely associations. The most common gating technique is ellipsoidal. The ellipsoidal shape is a consequence of the assumption that the error in the residual ($\tilde{z}_k$) is Gaussian [16]. The ellipsoidal gating defines a gate G such that the association is allowed if the norm of the residual vector (d^2) is within this gate G:

$$\begin{aligned} d_i^2 &= \tilde{\mathbf{z}}_i{}^T\mathbf{S}_i^{-1}\tilde{\mathbf{z}}_i \\ d^2 &= \sqrt{\textstyle\sum_{i=1}^{3} d_i^2} \leq G \end{aligned} \tag{3.10}$$

where $\tilde{\mathbf{z}}_i$ and $\mathbf{S}_i$ are the residual vector and its covariance matrix of the coordinates x, y or z, respectively, defined in (3.9). The norm d^2 is calculated for all combinations of measurements and tracks. When a measurement satisfies the gating inequality with a track, it is declared as a validated measurement for that track. Otherwise, it will be excluded from the possible assignments in the data association. This gating process will be revisited in the context-based privacy scheme, as explained in Chapter 7.

The gate size G can be calculated adaptively based on the probability of detection P_D and the residual vector. The probability of detection can be envisioned as the packet delivery ratio expected in vehicular network context. However, as stated in [23], d^2 is assumed to have Chi distribution χ^2_M where M is the degree of freedom or the dimension of the measurement vector. For the model specified in the previous section $(M = 6)$, G is set to be more than 19.

3.2.3 Data Association

After measurements are validated for each track, it is likely to have the same measurement in more than one gate, as illustrated in Figure 3.1. As it is not allowed to assign a measurement to multiple tracks, it is necessary to do association for all tracks simultaneously to avoid incorrect or sub-optimal solutions. There are several association approaches and they differ in how the assignment is accomplished. Some approaches, such as the global nearest neighbor (GNN), find the best measurement to update each track. However, there are others, such as joint probabilistic data association (JPDA), that incorporate several measurements with weighting probabilities to update a single track. Also, the assignment decision can be taken based on the measurements of the current scan or postponed several scans until finding the best hypothesis, as in multi-hypothesis tracking (MHT).

The GNN is the simplest data association approach as it handles the association problem in a straightforward way. It calculates a cost for each measurement-to-track assignment forming an assignment matrix. It uses an efficient method for solving the assignment problem to find the maximum number of possible assignments which minimizes the total cost. The cost function can be defined in multiple ways. For example, the cost function can be defined as the statistical distance of measurement j to track i as follows:

$$d^2_{G_{ij}} = d^2_{ij} + ln(|\mathbf{S}_{ij}|) \tag{3.11}$$

where d^2_{ij} is defined in (3.10) and $ln(|\mathbf{S}_{ij}|)$ is the logarithm of the determinant of the innovation covariance matrix $\mathbf{S}_{ij}$ defined in (3.9). This last term is used to penalize tracks with high uncertainty expressed in a large innovation matrix. There are several approaches that enhance the association of GNN such as branching to multiple hypotheses or calculating the cost function using subsequent scans. However, the GNN becomes obsolete because of the feasibility of more advanced techniques, such as JPDA and MHT [23].

The JPDA updates the track with a weighted average of all the measurements within its gate. The weighting function for assigning measurements to a track can be calculated as follows. For each scan, the probability of each hypothesis that assigns a validated measurement to a track is calculated. The probability of a particular measurement-to-track association is calculated as the sum of probabilities of all hypotheses that include this association. The JPDA is not appropriate for vehicle tracking because it results in a low tracking accuracy with closely spaced targets, as shown in [50]. Additionally, updating one track by multiple measurements is irrational, because it is guaranteed that different measurements or beacons necessarily correspond to different vehicles. Thus, updating a vehicle track by states of other vehicles results in deviation in the generated tracks. Finally, the complexity of JPDA is combinatorial because it requires generating all association hypotheses.

However, there is another simplified form of JPDA proposed in [50] which is referred to as nearest neighbor PDA (NNPDA). It aims to simplify the association calculations and avoid weighted-average updating feature in JPDA. It calculates a probability for each measurement to track association similar to JPDA, without generating the association hypotheses. It forms an assignment matrix with these probabilities and uses an assignment algorithm to select the optimal assignments. The probability P_{ij} of assigning a measurement j to track i is defined as:

$$P_{ij} = \frac{G_{ij}}{T_i + M_j - G_{ij}}, \; G_{ij} = \frac{e^{-d_{ij}^2/2}}{(2\pi)^{N_m/2}\sqrt{|\mathbf{S}_i|}} \tag{3.12}$$

where G_{ij} is the Gaussian likelihood function associated with the assignment of measurement j to track i, T_i is the sum of likelihood functions G_{ij} of track i and M_j is the sum of likelihood functions G_{ij} of measurement j. The d_{ij}^2 is the normalized distance between the measurement j and track i defined in (3.10) and the $|\mathbf{S}_i|$ is the determinant of the residual covariance matrix defined in (3.9). N_m is the dimension of the measurement vector. After calculating all probabilities, an assignment matrix is formed to obtain the optimal associations that maximize the sum of probabilities. This assignment problem is solved using an auction algorithm considering tracks as the bidders, beacons as the items and the bidding price as P_{ij}. We used an MATLAB implementation [112] of the auction algorithm proposed in [39]. These optimal associations are used to individually update each track by the associated beacon in the Kalman filter.

The MHT is different from GNN and PDA approaches in that it postpones the association decision for multiple subsequent scans. It generates hypotheses for all validated measurements with each track but it propagates (a subset of) them for subsequent time steps aiming to resolve the uncertainty. Since the propagation of hypotheses leads to combinatorial explosion, several techniques are used to reduce the complexity such as pruning, clustering or track merging.

The choice of the appropriate data association approach is crucial and depends on the application specifications and requirements. We used the NNPDA technique for the data association because it is more efficient than MHT, which enables real-time calculations even in a dense traffic. In addition, NNPDA achieves accurate association, as will be shown in Section 3.4.

3.2.4 Track Maintenance

A track maintenance phase is required to initiate, confirm and delete tracks. When a measurement is received and not assigned to a previously established track, a new track is initiated. However, this measurement may be a false alarm, thus this track is considered as a *tentative* track until it is confirmed in subsequent scans. The track confirmation can be typically done if M correlating measurements received in N scans and assigned to this track. Another approach is to define a score function for tentative tracks and confirm them once they exceed a predefined threshold. When a track is not updated for a while, it should be deleted to avoid further wrong associations and reduce the computational overhead. A typical deletion rule is to delete a track after a deletion tolerance interval of N consecutive scans with no update. Also, a score function can be used for this purpose.

In vehicle tracking, the track maintenance is simpler because lack of clutter. A track is initiated and confirmed immediately once a beacon is received and not assigned to a previously established track. For track deletion, the track is kept for a time-to-live T_{ttl} without an update; it is subsequently deleted. This T_{ttl} should be carefully handled with respect to the expected packet loss due to intentional (e.g., silent periods) or unintentional (e.g., channel congestion) reasons. If this parameter is small and several consecutive beacons are lost, the track will be rapidly deleted which will cause several discontinuities in the vehicle track. In contrast, if T_{ttl} is large, multiple vehicle traces may be merged into a single track.

3.3 Evaluation Metric

In the tracker evaluation, we used the *tracking percentage* as a metric for the tracker accuracy. To explain how this metric is calculated, we show first how the tracker practically works. Initially, it creates a set of tracks for beacons which appear in the first time step. Next, it assigns beacons of subsequent time steps to the established tracks or it may start new tracks. However, it may confuse and assign a beacon to a wrong vehicle track. Later, it may overcome this confusion and return assigning beacons to the original correct track. Therefore, the generated tracks must not be the same as the original traces due to these confusions. One track can be formed from several vehicle traces and one vehicle trace can be composed of multiple tracks at different times. We consider a successful tracker to be one that can produce *continuous* and *correct* tracks as long as possible. In the optimal case, the whole vehicle trace is assigned to a single track and each track is assigned to only one vehicle trace, resulting in 100% continuous tracking. To manage intermediate cases, we calculate the continuous tracking periods that a tracker can achieve for each vehicle trace. We then assign one track to only that vehicle trace that maximizes the length of the total tracking periods for all vehicles. Formally, the tracking metric can be defined as follows. Let $l(v,t)$ be the continuous tracking period when the vehicle trace v is assigned to the track t, $\forall v, t \in V, T$. τ_v is the maximum tracking period of v and obtained by solving the following assignment problem:

$$\begin{aligned}
\text{maximize} \; & \sum_{v \in V} \tau_v \\
\text{subject to} \; & \tau_v = \sum_{t \in T} l(v,t) \cdot a_{v,t}, \quad a_{v,t} \in \{0,1\}, \\
& \sum_{v \in V} a_{v,t} \leq 1 \quad \forall t \in T \quad \textit{and} \quad \sum_{t \in T} a_{v,t} \leq 1 \quad \forall v \in V.
\end{aligned}$$

This assignment problem can be solved using an auction algorithm considering tracks as the bidders, vehicle traces as the items and the bidding price as $l(v,t)$. Therefore, the tracking percentage can be defined as:

$$tracking\, percentage = \frac{\sum_{v \in V} \tau_v}{\sum_{v \in V} L(v)} \times 100 \tag{3.13}$$

where $L(v)$ is the lifetime of v. This metric is similar to the one used in [143] except that multiple vehicle traces can be assigned to the same track in different times and a single confusion is permitted in the tracking period.

3.4 Experiment Results

In evaluation, we use the VISSIM vehicle traces explained in Section 1.6.1. Since the position and velocity retrieved from VISSIM is accurately measured where it is not the case in reality. Thus, we add a normally distributed random noise, typically 1 m, to the position. Also, we assume vehicles obtain accurate speed

Table 3.1: Tracker evaluation parameters in urban and highway scenarios

	Urban		Highway	
Parameter	Range	Default	Range	Default
Arrival rate (Vehicle/hour)	100 - 600	300	300 - 1000	600
Desired speed (km/h)	30 - 70	50	80 - 130	100
Beaconing time t_b (s)	0.1 - 5	0.5	0.1 - 5	0.5
Position noise σ_p (m)	0 - 10	1	0 - 10	1
Speed noise σ_v (%)	0 - 10	2	0 - 10	2
Track time-to-live T_{ttl} (beacons)	1 - 10	2	1 - 10	2
Packet delivery ratio PDR	0.7 - 1	1	0.7 - 1	1
Simulation runs		10		

measurements from the wheel speed sensors used in Anti-lock Braking System (ABS). In typical conditions, the velocity noise can be maintained to be within 2% of the current speed [122]. The beaconing time t_b is assumed to be 0.5 s. Since the traces sampling interval is 0.1 s, we consider only one sample every five time steps to obtain the 0.5 s beaconing time. We run each experiment 10 times with different random noises. We evaluated a range of values for each of these parameters along with the traffic density and the desired speed offered in the simulation scenarios. The evaluated parameter ranges and their default values are shown in Table 3.1. In the next experiments, we show the effect of changing two parameters while assigning the remaining parameters to their default values. The error bars shown in the figures represent the standard deviation, if any.

3.4.1 Parameters Selection

Before discussing the experiment results of the tracker, parameters of Kalman filter and Gating should be adequately selected because they influence the tracking accuracy. We evaluated a wide range of parameter values repeatedly until the optimized value for each parameter is identified. Table 3.2 shows the test ranges for each parameter and its optimized value used in all experiments.

3.4.2 Anonymous Beacons

The tracker is evaluated using VISSIM vehicle traces described in Section 1.6.1. It is worthy to note that beacons are anonymized which means that the tracker uses only the spatiotemporal information (i.e., time, position and velocity) to

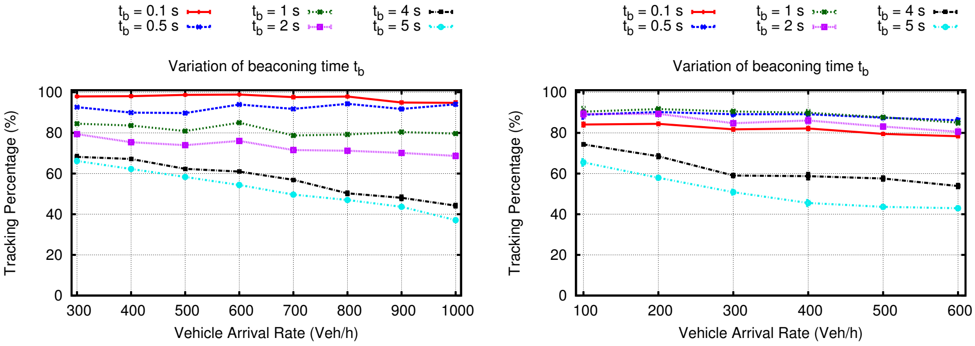
Table 3.2: Optimized values for Kalman filter and gating.

	Parameter	Test Range	Optimized Value
Kalman filter	p_0	20 - 70	40
	σ^2_{ap}	0.1 - 5	1
	σ^2_p	1 - 25	7
	σ^2_v	0.5 - 5	1
Gating	G	15 - 50	30

track vehicles. Our hypothesis is that if a tracker can achieve a high tracking accuracy in anonymous beacons, it will track vehicles more effectively when a pseudonym is used for several beacons.

The first experiment evaluates the tracker for different vehicle arrival rates with variant beaconing times t_b, as shown in Figure 3.3. In the highway scenario, the tracking percentage decreases with the increase of the length of the beaconing time t_b with a little effect of the arrival rate (for $t_b \leq 2$ s). This result is expected because the high speed of vehicles makes the change in their position faster which makes the tracker needs more frequent updates to achieve a higher accuracy. In contrast, in the urban scenario, the tracking percentages of $0.5 \leq t_b \leq 2$ s are almost similar regardless of the arrival rate. This small difference in the tracking percentages occurs because vehicles move near each other with small state changes and beaconing times up to 2 s are sufficient to track vehicles accurately. Also, the tracking percentage for $t_b = 0.1$ s is lower than those for $0.5 \leq t_b \leq 2$ s. This reduction occurs because frequent updates in a relatively low speed environment with the presence of noise makes the tracker confuse more among these nearby updates. In general, the beaconing times up to 1 s achieve a high tracking percentage of more than 80% in both scenarios. This finding emphasizes the trade-off between safety applications requirements of 10 Hz or even 1 Hz beacon rates and preventing tracking.

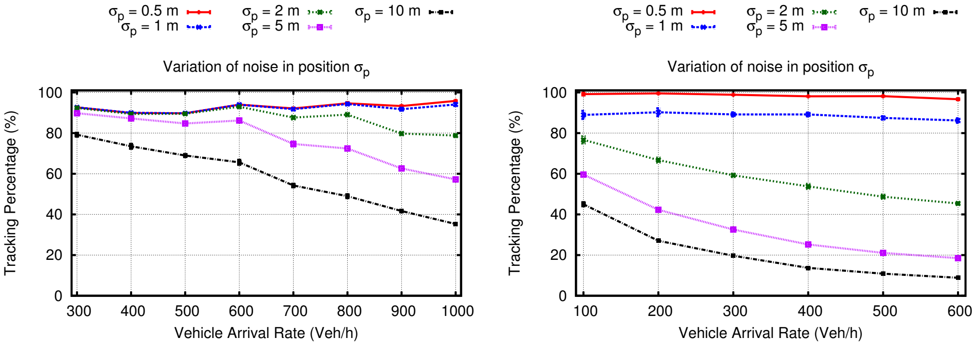
Next, we evaluate the tracker for different vehicle arrival rates with variant random noises in position, as shown in Figure 3.4. In both scenarios, the tracking percentage is more than 85% regardless of the arrival rate for less noisy positions ($\sigma_p \leq 1$ m). This result indicates that the positioning accuracy requirement of safety applications, such as lane change and forward collision applications, makes vehicles traceable, regardless of the vehicle density. Also, in intermediate arrival rates in the highway scenario (≤ 800 Veh/h), vehicles are still highly traceable (above 70%) even with largely noised positions ($\sigma_p \leq 5$ m). This result implies that noising the vehicle information is not sufficient to



(a) Highway scenario

(b) Urban scenario

Figure 3.3: Vehicle arrival rate versus beaconing time



(a) Highway scenario

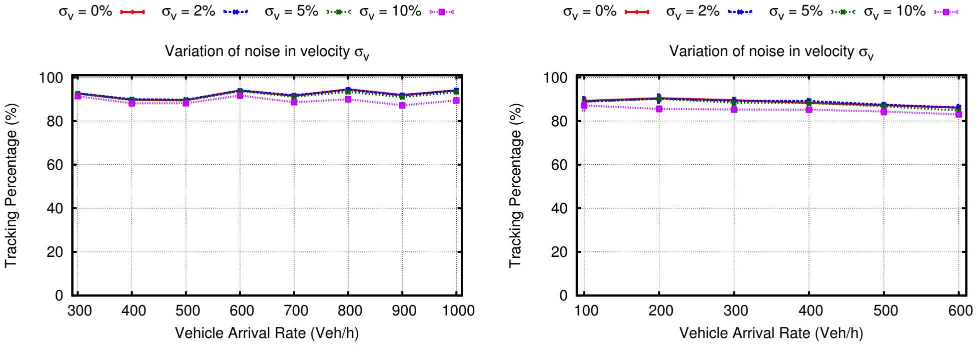
(b) Urban scenario

Figure 3.4: Vehicle arrival rate versus random noise in position

avoid tracking in a sparse traffic. In case of more noise in the urban scenario, the arrival rate becomes a factor and the tracker is more confused in linking beacons resulting in a lower tracking percentage. However, we can notice the impact of arrival rate in the urban scenario is greater than that in the highway scenario because the distances among vehicles are smaller.

Furthermore, the effect of noise in velocity is evaluated in Figure 3.5. Noises up to 5% of the current velocity achieve the same tracking percentage. Although larger noises in velocity (e.g., 10%) slightly reduces tracking, it has much lower effect than the noise in position.

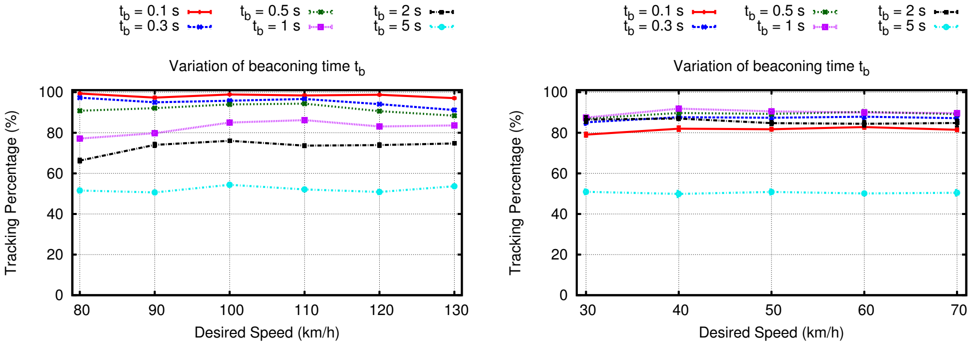
The next two experiments test the effect of the vehicle desired speed on the tracking percentage. Figure 3.6 shows the vehicle desired speed versus the beaconing time while Figure 3.7 presents the vehicle desired speed versus the



(a) Highway scenario

(b) Urban scenario

Figure 3.5: Vehicle arrival rate versus random noise in velocity



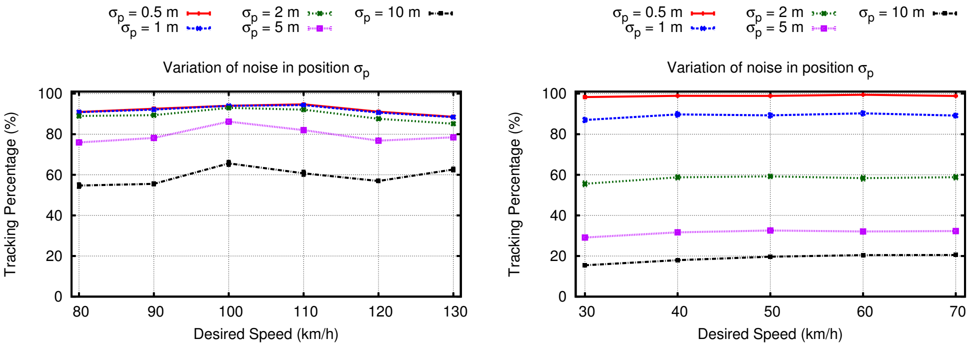
(a) Highway scenario

(b) Urban scenario

Figure 3.6: Vehicle desired speed versus beaconing time

random noise in position. In general, the desired speed has a slight effect on the tracking percentage in the highway scenario and almost no effect in the urban scenario. This behavior comes from that the desired speed does not change the actual traffic distribution or density in simulation so that it does not change the tracking percentage. The beaconing time and random noise in position produce the same tracking percentage as in their corresponding experiments with arrival rate.

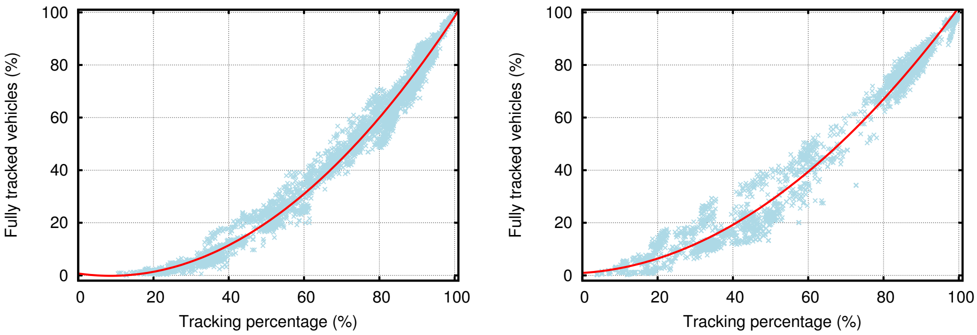
The tracking percentage metric represents the quality of tracking by showing how long the vehicle traces can be tracked. However, it does not show how many vehicles are completely tracked from start to end (i.e., $\tau_v = L(v)$). For example, the tracker can track on average 50% of the vehicle traces but in the same time there are many vehicles are completely tracked. Thus, we use an ad-



(a) Highway scenario

(b) Urban scenario

Figure 3.7: Vehicle desired speed versus random noise in position



(a) Highway scenario

(b) Urban scenario

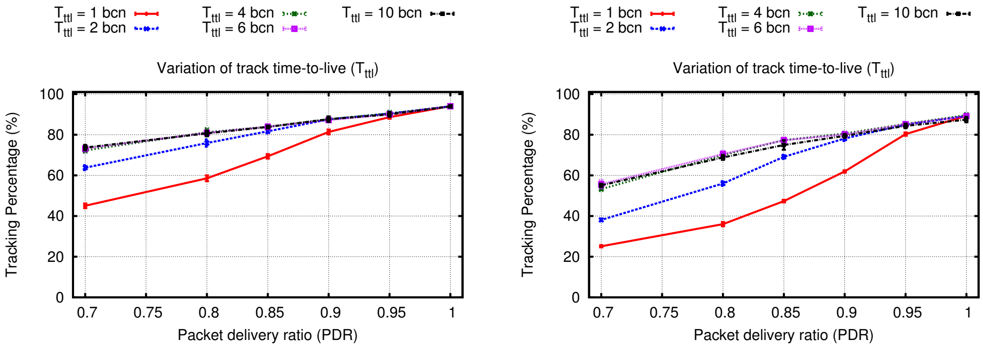
Figure 3.8: Completely tracked vehicles versus tracking percentage

ditional metric to clarify this case which is the percentage of vehicles that are completely tracked or so called *traceability*. Using thousands of simulation runs performed in the previous experiments, the relation between these metrics is illustrated in Figure 3.8. The samples of both metrics are fitted on a quadratic polynomial function drawn as red curves. These figures show that it is possible to completely track many vehicles, although the average tracking percentage is low. For example, 40% of vehicles can be completely tracked on average in the urban scenario when only a tracking percentage of 60% is achieved. Also, in both scenarios, at least 60% of vehicles are completely tracked on average for tracking percentage of 80%. This result indicates that even with conditions leading to intermediate tracking percentages, many vehicles can be completely tracked and totally losing their location privacy. Interestingly noted from Figure 3.8, the average of completely tracked vehicles is more in the urban scenario than the highway scenario.

3.4.3 Packet Delivery Ratio

In the previous experiments, it was assumed that the tracker is perfectly global so that it can eavesdrop every message broadcast to the network. However, this assumption is not realistic due to the typical limitations of wireless communication such as packet loss. Packet loss is common in wireless communication due to several reasons such as signal degradation and channel congestion. The effect of packet loss on vehicle tracking is that a random set of beacons is lost every time step and thus the tracker has incomplete knowledge about the traffic, which in turn reduces its tracking capability. We simulate the packet delivery ratio (PDR) by removing a random set of beacons of size equals to the loss ratio every time step. For example, to simulate a PDR of 0.8, we remove a one-fifth random set of beacons sent every time step. It may be not the best way to simulate the PDR because the packet loss is affected by more complex conditions in reality. However, we assume the tracking percentage will not differ significantly when the packet loss distribution is non-uniform over time.

In the first experiment, the PDR correlation with the track time-to-live (T_{ttl}) parameter is investigated. As discussed in Section 3.2.4, T_{ttl} affects the tracking tolerance against the loss of subsequent messages of a vehicle. A T_{ttl} of one time step means that the track is deleted if it is not updated for two consecutive time steps and so on. Both urban and highway scenarios are examined with a range of PDR between 0.7 and 1 along with several T_{ttl} values range from 1 to 10 beacons. We run simulations using the default values specified in Table 3.1.



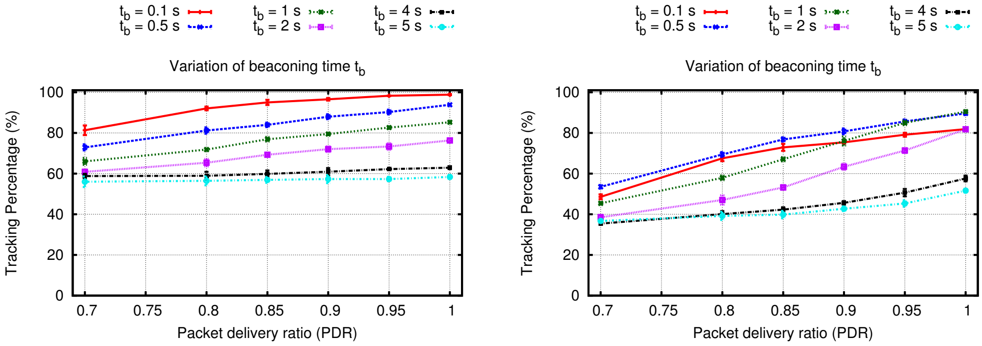
(a) Highway scenario

(b) Urban scenario

Figure 3.9: Packet delivery ratio versus track time-to-live (T_{ttl})

As shown in Figure 3.9, the track time-to-live (T_{ttl}) does not play any role in the case of the perfect packet delivery (i.e., PDR = 1). This is important as our previous results assumes a tolerance interval of two time steps and perfect PDR , thus, we do not need to repeat the previous experiments. However, for lower PDRs, the tracking percentage is significantly decreased but it can be improved by using non-short track time-to-live values (i.e., $T_{ttl} > 2$ beacons). However, longer T_{ttl} values ($T_{ttl} \geq 4$) do not enhance tracking already degraded by the packet loss. They achieve the same tracking accuracy. Thus, low values of the track time-to-live decrease the tracking percentage but the longer ones do not enhance it. Moreover, the tracking percentage is more degraded in the urban scenario than the highway scenario in lower PDRs (PDR ≤ 0.9). In the urban scenario, the tracker is more confused because the traffic is denser and the noise in position leads to wrong beacon associations.

Furthermore, we evaluate the effect of the packet delivery ratio with respect to the beaconing time, as shown in Figure 3.10. Based on the previous experiment, we choose the track time-to-live (T_{ttl}) to be 4 beacons. For the highway scenario, the tracking percentage is reduced linearly for short beaconing times ($t_b \leq 2$). However, the tracking percentage becomes almost constant for longer beaconing times regardless of the PDR . In the urban scenario, the tracking percentage decreases for all beaconing times with the decrease of the PDR . Thus, the reduction of the tracking percentage caused by the packet loss can be partially mitigated using short beaconing times.



(a) Highway scenario

(b) Urban scenario

Figure 3.10: Packet delivery ratio versus beaconing time ($T_{ttl} = 4$ beacons)

3.4.4 Beacon Information

After evaluating the proposed tracker, we evaluate the influence of the information contained in the beacon message on the tracking. Wiedersheim *et al.* [143] employed only time and position in their MHT tracker. However, their tracker accuracy is degraded significantly (up to 40%) for any random noise and beaconing times more than 1 s even with small traffic densities (75 vehicles and higher). Although the NNPDA is simpler than MHT, the NNPDA achieves a tracking percentage above 85% for position noises up to 1 m and above 70% for beaconing times up to 2 s according to the evaluated scenarios of different densities. These differences can arise from the tracking method, the simulation scenarios and/or the vehicle state model. In this section, we present the impact of the state model and the beacon information on the tracker accuracy. This experiment is crucial because it determines what is the necessary and sufficient information to be able to track vehicles effectively. It validates the assumption that the more information the tracker knows about vehicles, the more effective it can track them. To test the correctness of this hypothesis, we proposed two additional state models, the P and PVA models, similar to the model defined in Equations 3.3-3.6. The P model uses the vehicle position only in the measurement vector ($\mathbf{z}_k$) while the PVA model uses the position, velocity and acceleration. Note that the state vector ($\mathbf{x}_k$) of the P model includes velocity for better estimation results. The tracking percentage of both models are then compared with results obtained from our original model (i.e., PV model). The P Model is defined as follows:

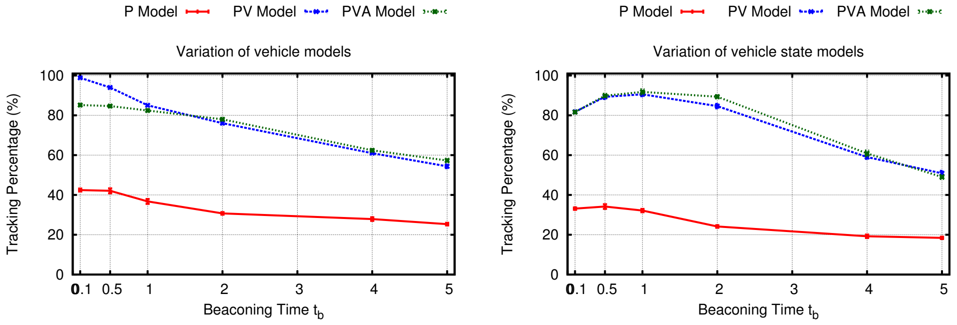
$$\mathbf{x}_k = \begin{bmatrix} p \\ s \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 1 & t_b \\ 0 & 1 \end{bmatrix}, \mathbf{z}_k = \begin{bmatrix} p \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{3.14}$$

$$\mathbf{Q} = \begin{bmatrix} t_b^4/4 & t_b^3/2 \\ t_b^3/2 & t_b^2 \end{bmatrix} \sigma_{ap}^2, \mathbf{R} = \begin{bmatrix} \sigma_p^2 \end{bmatrix} \tag{3.15}$$

While the PVA Model is defined as follows:

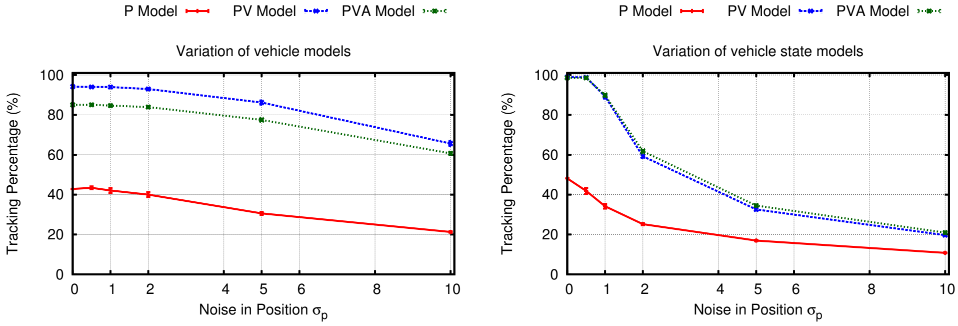
$$\mathbf{x}_k = \begin{bmatrix} p \\ s \\ a \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 1 & t_b & t_b^2/2 \\ 0 & 1 & t_b \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{z}_k = \begin{bmatrix} p \\ s \\ a \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.16}$$

$$\mathbf{Q} = \begin{bmatrix} t_b^4/4 & t_b^3/3 & t_b^2/2 \\ t_b^3/2 & t_b^2 & t_b \\ t_b^2/2 & t_b & 1 \end{bmatrix} \sigma_{ap}^2, \mathbf{R} = \begin{bmatrix} \sigma_p^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_a^2 \end{bmatrix} \tag{3.17}$$



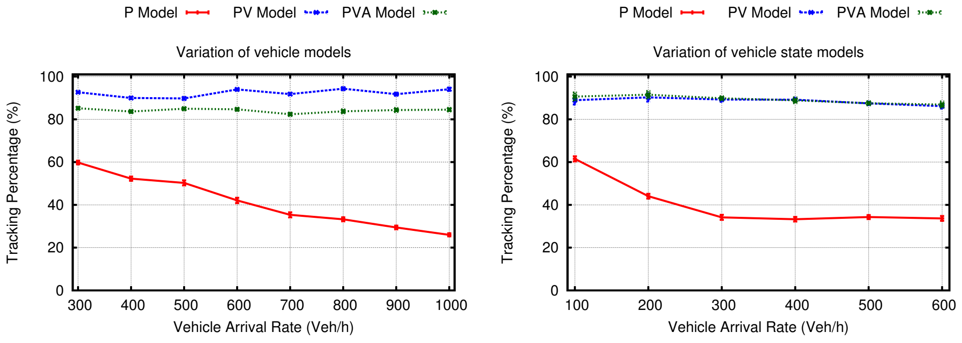
(a) Highway scenario (b) Urban scenario

Figure 3.11: Beaconing time versus vehicle state models



(a) Highway scenario (b) Urban scenario

Figure 3.12: Random noise in position versus vehicle state models



(a) Highway scenario (b) Urban scenario

Figure 3.13: Arrival rates versus vehicle state models

It is worthy to note that the previous matrices are for a single coordinate x, y, or z. We run the tracker using these models along with the original one (PV Model) on the highway and urban scenarios with similar parameters specified in Tables 3.1 and 3.2. As shown in Figures 3.11, 3.12 and 3.13, the P model performs worse than the other models for different arrival rates, position noises and beaconing times. This confirms that the degradation in the tracking accuracy in [143] essentially caused by the employed model. Thus, position information is not sufficient to achieve a reliable vehicle tracking. Additionally, employing the position and velocity information is sufficient for vehicle tracking and provides similar tracking accuracy to employing acceleration in addition.

3.4.5 Comparison with MHT Tracker

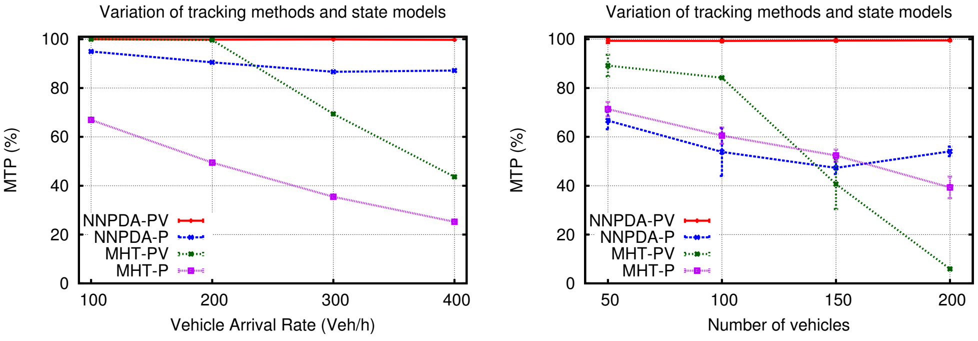
The presented results in previous sections show reasonable effectiveness of the proposed NNPDA tracker. However, it is important to confirm its robustness by comparing with other trackers based on advanced data association algorithms such as MHT. The MHT tries multiple hypotheses over subsequent time steps rather than taking an assignment decision based on the information of the current time step. Also, it is desirable to apply the tracker on different vehicle trace datasets to confirm its generality. Therefore, we obtained the MHT tracker and the traces dataset from Wiedersheim *et al.* [143]. Their tracker uses a vehicle state model based on positions only which is not sufficient for effective tracking, as explained in Section 3.4.4. Hence, we modified their MHT tracker to consider both position and velocity in the state estimation, as defined in Equations 3.3-3.6. Also, they calculate the tracking period differently because they allow a single track to be assigned to more than one vehicle trace in different times. We adopted their calculation method but used the *mean tracking percentage* (MTP) as the comparison metric which can be defined as follows:

$$MTP = \frac{\sum_{v \in V} \max_{t \in T} l(v,t)}{\sum_{v \in V} L(v)} \times 100 \tag{3.18}$$

where $l(v,t), \forall v, t \in V, T$ is the continuous tracking period when the vehicle trace v is assigned to the track t and $L(v)$ is the lifetime of v. Finally, they used their STRAW traces explained in Section 1.6.2 in the tracker evaluation. It is clear that the STRAW scenario has much more intersections and road segments than that in VISSIM scenarios. However, the VISSIM provides more realistic traces because it uses a car-following model that considers physical and psychological aspects of the drivers. To take advantage of both traces, we use the STRAW traces and the urban scenario of VISSIM .

In Figures 3.14, we show the MTP of noiseless positions as obtained from the traces dataset versus the vehicle arrival rate or density. We notice that the

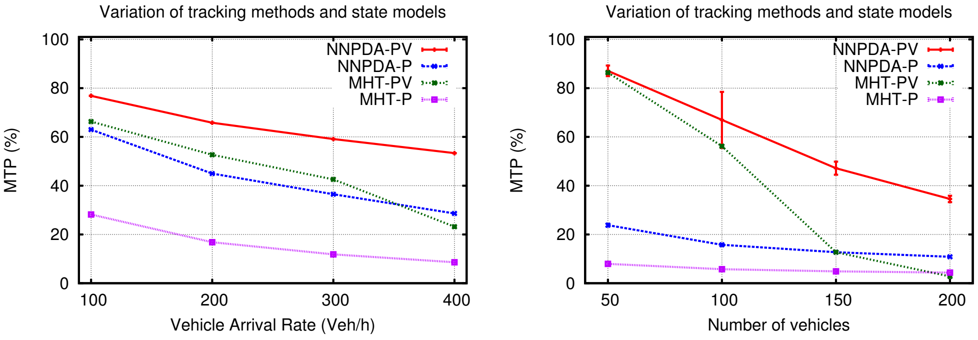
NNPDA with the PV model achieves perfect tracking in all densities and scenarios. The MHT with PV model achieves a high accuracy in the intermediate vehicle densities, otherwise its MTP is reduced even to the lowest of all other variations in the STRAW scenario. The NNPDA with P model is not stable; its MTP is high in the VISSIM scenario while it is low in the STRAW scenario. Lastly, the MTP of the MHT with P model is low and reduces with the vehicle density.



(a) VISSIM urban Scenario

(b) STRAW traces

Figure 3.14: Comparison of tracking methods and vehicle state models in noiseless positions.



(a) VISSIM urban Scenario

(b) STRAW traces

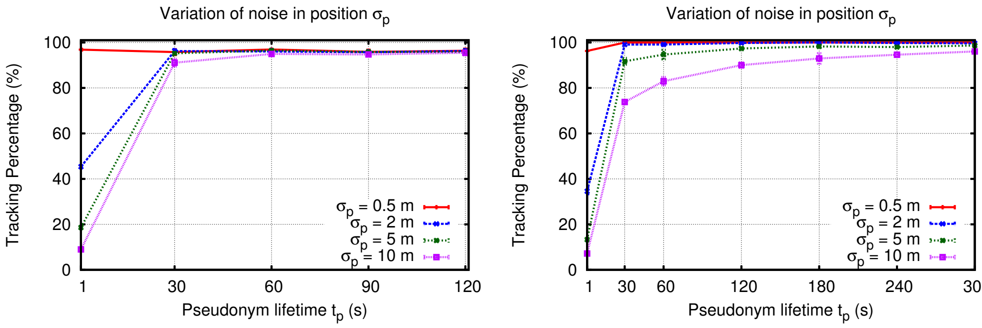
Figure 3.15: Comparison of tracking methods and vehicle state models in noisy positions of 2 m.

In Figure 3.15, different tracking methods and model are evaluated for noisy positions of 2 m in both scenarios. The NNPDA with PV model achieves the highest MTP among the others on average. The MHT with PV model achieves a comparable percentage in the intermediate densities, although its MTP decreases more in the higher densities. The MHT and NNPDA of P model are achieves low MTP in the presence of noise, although the NNPDA-P performs better in the VISSIM scenario. According to these results, the robustness of our NNPDA tracker in comparison with the MHT tracker [143] is confirmed. We will use the NNPDA tracker in evaluating privacy schemes as a global adversary and enhancing the quality of safety applications by embedding it inside vehicles, as will be explained in the next chapters.

3.4.6 Pseudonymous Beacons

When a vehicle uses the same pseudonym for several beacons, the tracker can easily correlate these beacons. The tracker assigns beacons to the tracks by matching similar pseudonyms. The tracker uses the data association algorithm (i.e., NNPDA) only when correlating beacons of new pseudonyms with unmatched tracks. The tracker keeps all encountered pseudonyms in a list for the pseudonym maximum lifetime defined by the privacy scheme. After that time, the pseudonym is removed from this list. A pseudonym is identified as new if it does not exist in this pseudonyms list.

To give an illustration of how the tracking percentage can be enhanced with periodically-changed pseudonyms, we evaluate scenarios of 0.5, 5 and 10 m normally-distributed position noises with pseudonyms changed every fixed time t_p ranging from 0.5 to 300 s. To avoid synchronization effect and ensure that a vehicle changes its pseudonym at least once, vehicles are forced to change its pseudonym within the first 10 time steps of its arrival. Also, it is worthy to note that the tracker does not exploit the knowledge of that t_p has a fixed length in order to predict when exactly a vehicle change its pseudonym. The fixed t_p may harden the tracking vulnerability because it increases the number of vehicles that change their pseudonyms simultaneously. Since the average lifetime of vehicles is relatively short in the VISSIM scenarios, we also evaluated the STRAW vehicle traces. Figures 3.16 illustrate the tracking percentages versus the pseudonym lifetime t_p with different noises. We selected the highest vehicle density from each dataset which harden the tracking mission. For the VISSIM dataset, the urban scenario of 600 vehicle/hour arrival rate is selected while, for the STRAW dataset, the vehicle density of 200 vehicles is chosen. We can notice that a tracking percentage of more than 90% and 70% can be achieved even with the presence of noise of 10 m (i.e., the common GPS noise) when pseudonyms are changed every 30 s for the VISSIM and STRAW datasets, respectively. Higher tracking percentages are attainable with longer pseudonym lifetimes. We show also the tracking percentage in the theoretical case where a vehicle uses a new pseudonym every beacon. The tracking per-



(a) VISSIM Scenario (600 Veh/h)

(b) STRAW Scenario (200 Veh)

Figure 3.16: Tracking with pseudonyms in the highest vehicle density of the VISSIM urban scenario and the STRAW traces.

centage is dropped to about 8% in both datasets when the noise in position is 10 m. These results confirm three important findings. First, using pseudonyms for several beacons increases the tracking vulnerability significantly even if they are changed every relatively short periods (e.g., 30 s) and the positioning noise is large (i.e., $\sigma_p = 10$ m). Second, changing pseudonyms frequently does not reduce tracking vulnerability when small noises in position are expected (e.g., $\sigma_p = 0.5$ m) even in dense traffic, as shown in Figures 3.16. Third, simultaneous pseudonym changes among nearby vehicles is desirable to confuse the tracker but it is not sufficient to avoid tracking. The most frequent and simultaneous pseudonym change occurs when it is changed every beacon. Based on these results, vehicles are traceable with a very high likelihood specially when accurate measurements are used in the beacon messages.

3.4.7 Tracking with Silent Period

When the vehicle traces include a random silent period before a pseudonym change, the tracker is tuned to handle this expected silence. The tracker basically holds a vehicle track without update till track time-to-live (T_{ttl}) time steps and deletes it after that time. We added an extra parameter of the maximum silence period (*max-silence*) that can be employed by a privacy scheme. The tuned tracker only marks a vehicle track as inactive after T_{ttl} time steps and holds it for additional *max-silence* time steps. When the tracker assigns beacons of unmatched pseudonyms to its current tracks list, it only considers inactive tracks. This modification increases the tracking percentage since it eliminates matching beacons of new pseudonyms with unrelated tracks.

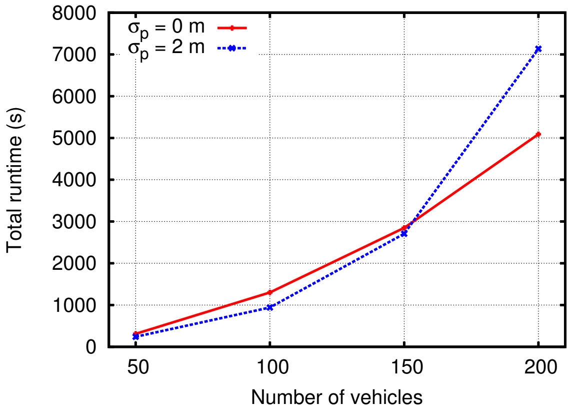


Figure 3.17: Runtime of the vehicle tracker using anonymous beacons of the STRAW traces

3.5 Tracker Complexity and Efficiency

The complexity of the vehicle tracker is $O(KVN)$ where V and N are the number of beacons and tracks per time step, respectively, and K is the total number of time steps. Generally, $N \simeq V$, but when there are many confusions and the tracker creates many new tracks for unmatched beacons, then $N \gg V$. The standard implementation of the Kalman filter requires $O(d^3)$ [107], where d denotes the dimension of the vehicle state, because of the matrix inversion and multiplication operations. Since d is constant, we assume the complexity of Kalman filter is constant.

We implemented the tracker using MATLAB and run our experiments on an Intel QuadCore i7-4800MQ @ 2.70GHz CPU. We calculate the total running time required to track anonymous beacons of the whole STRAW traces of different densities, as shown in Figure 3.17. We observe an exponential runtime with the increase of the vehicle density. The exponential rate increases faster with the presence of position noise due to the increase of tracker confusions and creation of false tracks every time step. Moreover, we notice that the tracker can process the whole traces of 2000 time steps in about 1000 seconds with a vehicle density of 100 vehicles. This means that the tracker can track anonymous beacons in real-time with intermediate vehicle densities. In the pseudonymous beacons, the runtime is dramatically decreased even with short pseudonym lifetimes. For example, the tracker can process the whole densest scenario of 200 vehicles with a short pseudonym lifetime of 30 s in less than 100 s regardless of the position noise.

3.6 Tracker Enhancements

Although the proposed NNPDA tracker achieves a reasonable accuracy in tracking vehicles in different traces and conditions, this accuracy can be further enhanced. First, beacons contain additional static data, such as vehicle type and size. If this information is additionally used in tracking, it will help in discriminating between confusing beacons.

Second, the road geometry can be exploited in the tracking algorithm itself. There are several ground target tracking algorithms that use the road map and its geometry to predict the vehicle state more realistically such as [125, 148, 134]. Road curvature and surface, velocity limit and road direction are examples of the constraints that can be imposed to the state estimation. These constraints lead to better estimations which in turn lead to better data association and tracking accuracy.

3.7 Summary

In this chapter, the multi-target vehicle tracking is thoroughly discussed and a vehicle tracker based on the NNPDA algorithm is proposed and evaluated using different vehicle traces datasets. The experiment results can be summarized as follows:

- Anonymous beacon messages can be effectively and accurately tracked (tracking percentage more than 80%) for beaconing times up to 1 s and position noises up to 1 m in both urban and highway scenarios and regardless of the vehicle density.

- A reasonable number of vehicle traces can be entirely tracked from anonymous beacon messages even in conditions leading to intermediate tracking percentages. For example, 30% of traces can be completely tracked when the tracking percentage is only 60%.

- Low packet delivery ratios (PDR) reduce the tracking accuracy but this reduction can be mitigated by short beaconing time of 0.5 s or shorter. A tracking percentage of 80% can be achieved even with a PDR of 0.85.

- The position and velocity are the sufficient and necessary information to effectively track anonymous beacon messages.

- The proposed tracker and vehicle model achieved higher tracking accuracy than the MHT tracker in both noiseless and noisy vehicle traces.

- Beacon messages identified by periodically-changed pseudonyms can be tracked more effectively, even with a large position noise up to 10 m. This result confirms the need for additional mechanisms to prevent tracking more than the periodical pseudonym change.

Based on these results, the trade-off between the safety application requirements and location privacy is clearly highlighted. Safety applications require beaconing time up to 1 s and position noise up to 1 m which are sufficient for accurate and continuous vehicle tracking. This finding asserts the need for protecting the driver's privacy by preventing vehicle tracking without hindering the operations of safety applications.

4 Measuring Location Privacy

4.1 Introduction

Westin defined privacy as that “the right to control, edit, manage, and delete information about them[selves] and decide when, how, and to what extent information is communicated to others” [141]. Location privacy is a special type of privacy which concerns the individual location. Location privacy is studied in different application areas such as databases, location-based services and mobile networks. Although there is a large number of privacy mechanisms proposed for VANET, there is a lack of consensus on suitable privacy metrics [105]. Each proposed privacy scheme is evaluated using a different metric which makes comparing the effectiveness of different schemes difficult.

Privacy is related to other concepts, such as anonymity, untraceability, unlinkability, unobservability and pseudonymity, which are essential for understanding and measuring privacy [29, 106]. We briefly explain these concepts before discussing the privacy metrics. *Anonymity* of a subject means that the subject is not identifiable within a set of subjects, the anonymity set [106]. For example, the sender of a message is anonymous when it cannot be identified who sent this message. According to this definition, anonymity is more than hiding or eliminating the identity of an action (e.g., removing the sender address from a message) because the identity can be guessed using other information sources or previous knowledge. For example, if a message was sent from a workplace on the weekend and it was known to the attacker that only Alice was at work in that day, in this case, the sender of this message could be easily re-identified even if the message is apparently anonymous. The anonymity definition states this condition by relating the anonymity of a subject to other subjects that may perform the action (i.e., anonymity set). If the anonymity set equals to one or the subject has unique characteristics from other members of the anonymity set, the subject is not anonymous.

Untraceability concerns making it difficult to correlate different actions performed by the same subject together [29]. Anonymity is necessary but not sufficient to guarantee untraceability. Subsequent actions can be individually anonymous but the adversary can use similar attributes of these actions to correlate them. *Unlinkability* usually generalizes the anonymity and untraceability concepts [29]. Unlinkability of two or more items of interest (e.g., subjects, messages and actions) from an attacker's perspective means that it cannot sufficiently distinguish whether these items are related or not, as defined in [106]. Therefore, anonymity means the subject and its actions are unlinkable and untraceability means actions of the same subject are unlinkable. In contrast to anonymity, *unobservability* concerns hiding the item itself instead of the identity. Unobservability includes the adversary unawareness about the action and the anonymity of the subject of that action [106]. *Pseudonymity* means that using a pseudonym instead of a real identity to identify oneself [29]. A pseudonym is an identifier of a subject other than and unlinkable to one of the subject's real identities. If a subject is using multiple pseudonyms, it is important that they are unlinkable to ensure untraceability.

Measuring location privacy of a privacy scheme requires quantifying its ability to fulfill the requirements of each concept against a well-defined adversary. We consider the adversary model defined in Section 1.5.2 which aims to reconstruct vehicle traces from their beacon messages. Therefore, we do not measure the pseudonymity in the metric because it should be fulfilled in the pseudonym issuing process which is out of the scope of this thesis. But we assume that vehicles use pseudonyms obtained from a service provider and use one pseudonym at a time. Unobservability is also ignored because it is assumed for safety applications to frequently broadcast the vehicle state unencrypted. Therefore, the privacy metrics presented and proposed in this chapter concern only anonymity, untraceability and unlinkability.

In this chapter, an overview of the existing privacy metrics for VANET is presented. We then propose the distortion metric that is used throughout the thesis in evaluating and comparing privacy schemes. An experimental comparison among the discussed metrics is explained at the end of the chapter.

4.2 Privacy Metrics

4.2.1 Anonymity Set Size

Vehicles are assumed to broadcast beacon messages continuously with their pseudonym, position, speed and heading. To provide anonymity and unlinkability, they are changing their pseudonyms periodically. Therefore, the basic location privacy metric is to measure the *anonymity set size*. The anonymity set of a target vehicle is the vehicles in which this target vehicle is not identifiable or distinguishable with respect to its location. For example, an anonymity set may be formed when two or more nearby vehicles change their pseudonyms in the same time. In this case, the adversary may confuse about the actual location of the target vehicle since it may be any vehicle from the anonymity set. A closely related metric is the *k-anonymity* which basically refers to an anonymity set with a minimum size k, where the target is indistinguishable from at least $k-1$ vehicles.

One shortcoming of this metric is that it is not necessary for all members of the anonymity set to be equally likely the target vehicle, from the adversary perspective. The adversary can calculate probability distribution for the anonymity set based on the spatiotemporal information in beacons so that less-likely correlations can be excluded from the anonymity set. Therefore, the anonymity set size is not a suitable location privacy metric because it cannot deal with nonuniform probability distributions of the anonymity set [40, 121]. Despite its unsuitability, the anonymity set size is used in some recent works, especially in analytical approaches, such as [87, 99].

4.2.2 Entropy

To handle the shortcomings of the anonymity set size, Serjantov and Danezis [121] and Díaz *et al.* [40] proposed an information theoretic metric, the Entropy, to measure the anonymity. Let $\mathcal{A}$ represent the anonymity set and p_i is the probability assigned by the adversary for each member in $\mathcal{A}$ to be the target such that $\sum_{i=1}^{|\mathcal{A}|} p_i = 1$, then the entropy $\mathcal{H}$ can be defined as:

$$\mathcal{H} = -\sum_{i=1}^{|\mathcal{A}|} p_i \cdot \log p_i \tag{4.1}$$

According to this definition, the entropy of a vehicle equals to zero while the same pseudonym is used for several beacons. Upon a pseudonym change, the entropy is calculated based on the probability distribution assigned by the adversary. The entropy achieves its maximum value when the probability distribution is uniform (i.e., $\mathcal{H}_{max} = \log_2 |\mathcal{A}|$). It decreases in other distributions till it reaches zero when only one p_i equals one and the rest equals zero. Since $\mathcal{H}$ is unbounded, Díaz *et al.* [40] proposed an extended metric, the *normalized entropy* $\mathcal{H}_n$, to measure the degree of anonymity:

$$\mathcal{H}_n = \frac{\mathcal{H}}{\mathcal{H}_{max}} \tag{4.2}$$

The entropy expresses on the adversary *uncertainty* about the linkability of a new pseudonym to the target vehicle. The given definitions measure the entropy of a single *mix* which is formed by simultaneous pseudonym changes of several vehicles. To calculate the overall entropy of a vehicle trace, entropies of individual mixes that occurred in the whole trace are summed together assuming entering consecutive mixes is independent, as presented in [72, 53].

Alternatively, the average, minimum and maximum can be calculated over all mixes to provide the expected, lower-bound and upper-bound of the adversary uncertainty, respectively, as adopted in [21, 73].

The entropy is intensively used in evaluating location privacy schemes in mobile and vehicular networks. However, the method that calculates the anonymity set and their probabilities differs from a scheme to another based on the assumed system and adversary models. Beresford and Stajano [19] used entropy to evaluate the anonymity of a mix-zone placed in specific locations. When users enter a mix-zone, they change their pseudonyms and exit the mix-zone after an unknown period of time (guaranteed by the mix-zone shape). The adversary calculates all possible mappings between old and new pseudonyms based on a movement probability matrix. This matrix is estimated based on the adversary knowledge of the source/destination frequencies. The normalized probabilities of these mappings are used in calculating the entropy. Later, Buttyán *et al.* [27] introduced the mix-zone concept into VANET and used the entropy to identify the *effective size* of the anonymity set. The apparent size of the anonymity set is the number of vehicles that exit a mix zone in the observed period. However, they showed that the effective size is much less due to the non-uniformity of the probability distribution. For the entropy probabilities, they calculate a probability p_{jt} for each exit event which is given by $p_{jt} = q_{sj} f_{sj}(t)$. q_{sj} is the probability that the vehicle chooses port j as its exit port given that it entered the mix zone at port s and $f_{sj}(t)$ is the probability that the vehicle covers the distance between ports s and j in time t. This probability calculation is similarly used by Freudiger *et al.* in [53] for evaluating their Cryptographic MIX-zones (CMIX) protocol.

Sampigethaya *et al.* [114] proposed the silent period as a type of dynamic mix-zones where vehicles keep silent for a random period before changing their pseudonyms. The anonymity set is considered to be all vehicles that update their pseudonyms in the reachable area of the target during the silent time range. The reachable area is calculated based on the target speed range, the road restrictions and the minimum and maximum silent time specified by the scheme. They calculated the probabilities of the anonymity set based on two different tracking methods: simple and correlation tracking. In the simple tracking method, all vehicles are assigned an equal probability. In the correlation tracking, the estimated location of the target vehicle is calculated based on its last known location, speed and direction at every time step during silence. The obtained location estimations are compared with the locations of other vehicles in the anonymity set. The adversary calculates non-uniform probability distribution based on the proximity between the vehicle locations and the corresponding estimated target location.

Although the popularity of entropy metric, it has several shortcomings pointed out in the literature. Tóth *et al.* [133] showed that a high value of entropy may not mean high anonymity especially when there are many low probable mappings that can be ignored by the adversary. Additionally, an entropy threshold cannot be specified for mix-zones so that the anonymity is confirmed if the estimated entropy is greater than this threshold value. Palanisamy and Liu [95] proposed to use the *pairwise entropy* which measures the deviation of the mapping probabilities in a pairwise fashion. The pairwise entropy between two users i and j is the entropy obtained by assuming that users i and j are the only members of the anonymity set. If the pairwise entropies $\mathcal{H}(i, j)$ and $\mathcal{H}(j, i)$ when i exits as i' and j exits as j' are both close to 1, it means that the attacker is highly uncertain about this mix.

Apart from the calculation details, Fischer *et al.* [49] argued that entropy-based metrics are not suitable to measure unlinkability because they do not distinguish among different probability distributions of linking subsequent messages estimated by different attackers. Moreover, Shokri *et al.* [123] claimed that the entropy and, of course, the anonymity set size metrics are not suitable for quantifying location privacy. The entropy shows how uniform versus condensed the estimated distribution and, in consequence, how certain the adversary about his decision. The higher the entropy becomes, the lower the adversary's certainty. However, the entropy does not derive any clue about the correctness of this decision. It may happen that the adversary is certain about his estimate with a high probability but, at the same time, this estimate is largely different from the actual user's location. This occurs because of the limitation and incompleteness of the adversary's knowledge about the actual situation.

4.2.3 Traceability

Another approach for measuring the location privacy is to calculate how long an adversary can track vehicles. Tracking vehicles or linking segments of different pseudonyms is inversely proportional to the location privacy. Identifying user trajectories and movement patterns is an essential step for privacy breaches (i.e., re-identification and localization attacks) [69].

There are several approaches to measure traceability. Huang *et al.* [71, 72] measured how long a node can be tracked continuously in evaluation of silent period schemes in mobile networks. They used the terminology of Maximum Tracking Round (MTR) which is the number of identifier rounds that a node is tracked continuously after its first identifier update. Consequently, the *maximum tracking time* is the MTR multiplied by the lifetime of the identifier. Sampigethaya *et al.* [114] defined the maximum tracking time differently as the maximum cumulative time that the target anonymity set size remains as one. Similarly, Hoh *et al.* [69] proposed the time-to-confusion metric which is the tracking time until the adversary uncertainty (i.e., entropy rather than the anonymity set) rises above a preset threshold. Also, they proposed another similar metric based on distance rather time in [68] which is called distance-to-confusion.

In the context of fixed mix-zones at road intersections, Buttyán *et al.* [27] and Freudiger *et al.* [53] evaluated mix-zones by the success probability of an adversary to track vehicles. This success probability is calculated by the ratio of the number of successfully mapped vehicles to the total number of vehicles in a mix-zone, averaged over all mix-zones. Furthermore, Buttyán *et al.* [28] used the spatiotemporal information in every two beacons to calculate the acceleration of the vehicles to accurately predict the next position. Then, they measure the tracking success rate which tracked vehicles from their departure to their destination. Wiedersheim *et al.* [143] measured the traceability as the average duration of each correctly tracked vehicle. However, they allow for the reconstructed traces to include false samples from traces of other vehicles. We used traceability in [44, 45, 46] with two different definitions. First, it is measured by the tracking percentage as defined in Equation 3.13. Second, it is measured by the percentage of vehicle traces whose a tracking percentage more than a preset threshold (e.g., 95%). Our calculation methods for traceability will be discussed in detail in Section 4.3.

4.2.4 Distortion

The last approach for measuring location privacy is to calculate the error or distortion of the reconstructed tracks compared to the actual traces. Hereafter, a *trace* refers to the original vehicle trace and a *track* refers to the reconstructed trace by the adversary. Hoh and Gruteser [67] proposed the expected distance error, which captures the adversary accuracy in estimating a user position. They defined the expected distance error for a path as:

$$E[d] = \frac{1}{NK} \sum_{k=1}^{K} \sum_{i=1}^{I} p_i(k) d_i(k) \tag{4.3}$$

where d_i represents the total distance error between the correct hypothesis and the hypothesis i for all user locations at a time step k. p_i is the probability of the hypothesis i obtained from the MHT algorithm used in reconstructing the user paths from positions sent every time step. N is the number of users and K is the total time steps. Similarly, Shokri *et al.* [124] defined an expected distortion metric which can be calculated as follows. First, they find the latest position from a user observed at or before a time step t, which is denoted by e_t. Then, all paths that start from e_t and end at t are identified to calculate

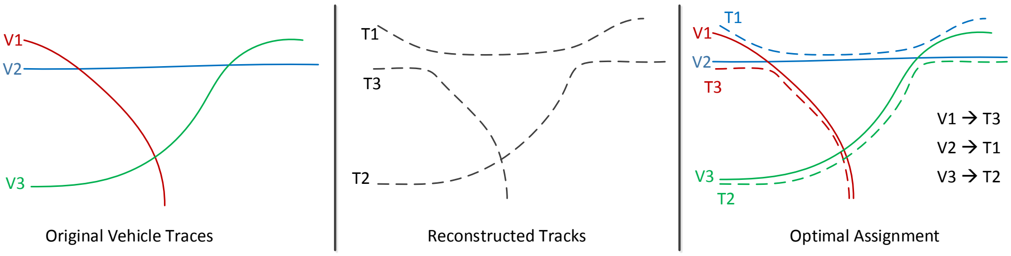


Figure 4.1: Traceability metric illustration

the expected user positions and their corresponding probabilities. At last, the expected distortion at time step t is the total weighted distance between the expected positions and the actual position multiplied by their corresponding probabilities. They also defined the distortion-based traceability which is the tracking time until the distortion exceeds a preset threshold.

4.3 Proposed Location Privacy Metric

Based on the presented metrics and criteria, we adopt a combined metric that is based on the traceability and distortion. It is important to measure both aspects to determine how long the adversary can track a vehicle and how different the reconstructed tracks from the actual traces. We hypothesize that reconstructing the entire vehicle trace is necessary to breach the driver privacy. This hypothesis is inferred from research works of re-identifying anonymous traces which use work/home location pairs [59], top N locations [152] or geosocial networks [33]. All these works depend on finding the frequently visited places of the user over a long time (e.g., several weeks). In VANET, these places can be identified by correlating the source and destination of each trip, which necessitates the ability of reconstructing the entire user traces. If the adversary is unable to reconstruct complete traces, then clustering techniques used in the re-identification process will fail in finding the driver places.

We investigate traceability more thoroughly since comparing the reconstructed tracks with the original vehicle traces is not trivial, as illustrated in Figure 4.1. In this example, there are three traces V1, V2 and V3 (drawn as solid lines on the left) that are reconstructed into three tracks T1, T2 and T3 (drawn as dashed lines on the middle). By visually comparing both sets, it is clear that each track is reconstructed from partial segments of the original traces. For example, T1 is reconstructed from segments of V1, V2 and V3. Traceability metrics presented previously in this chapter may fail to reflect the actual traceability level of this adversary. The main issue of their operation is that they assign tracks to vehicle traces during the tracking process. In other words, they assume the track firstly assigned to a vehicle trace should continue with this trace till its end as in [67, 71, 114]. However, this early assignment underestimates the length of the reconstructed tracks. For example, if the traceability of V1 is measured by assigning T1 to V1, then this metric shows a very short tracking time, although V1 is reasonably reconstructed by T3. Therefore, it will be more effective if tracks are assigned to the vehicle traces globally after the tracking process is complete. The track-to-trace assignment is basically a nonlinear *assignment problem* where the total benefit should be maximized. The benefit represents the tracking period when a track t^1 assigned to a vehicle trace v continuously. The assignment program is previously presented in Section 3.3 but that metric focuses on measuring the tracker capability, and thus, it considers only how long vehicles are continuously tracked. However, a traceability metric for privacy should reflect how this tracking capability threatens the user privacy. For example, a traceability-based privacy metric may correlate the tracking percentage to the probability of re-identifying anonymous tracks.

Therefore, the assignment program is re-discussed here to highlight this difference. Let $l(v,t), \forall v,t \in V,T$ be the maximum continuous tracking period when the track t is assigned to the vehicle trace v. Note that t can be assigned to v for disconnected segments at different times. In this case, $l(v,t)$ represents the longest segment. The disconnected segments are not summed together because the tracking is discontinued and the track may be assigned to another vehicle trace during this discontinuity. The adversary cannot reconnect these segments and filter out this wrong assignment period because the adversary does not know if he is confused or not. Let τ_v be the maximal tracking period of v; and it can be obtained by solving the following assignment problem:

$$\begin{aligned} \text{maximize } & \sum_{v\in V} \tau_v \\ \text{subject to } & \tau_v = \sum_{t\in T} l(v,t)\cdot a_{v,t}, \quad a_{v,t}\in\{0,1\}, \\ & \sum_{v\in V} a_{v,t} \le 1 \quad \forall t\in T \quad \mathit{and} \quad \sum_{t\in T} a_{v,t} \le 1 \quad \forall v \in V. \end{aligned} \tag{4.4}$$

Here, $a_{v,t}$ is the assignment function which equals one if the track t should be assigned to the vehicle trace v and equals zero otherwise. Note that not all tracks must be assigned to a vehicle trace because the number of tracks can be greater than the number of vehicle traces as some tracks are reconstructed from partial vehicle traces. Also, not all vehicle traces must be assigned to a track because its $l(v,t)$ may not contribute to the maximal $\sum_{v\in V}\tau_v$. In this case, τ_v

[1] In the rest of this chapter, t refers to a track rather than a time step.

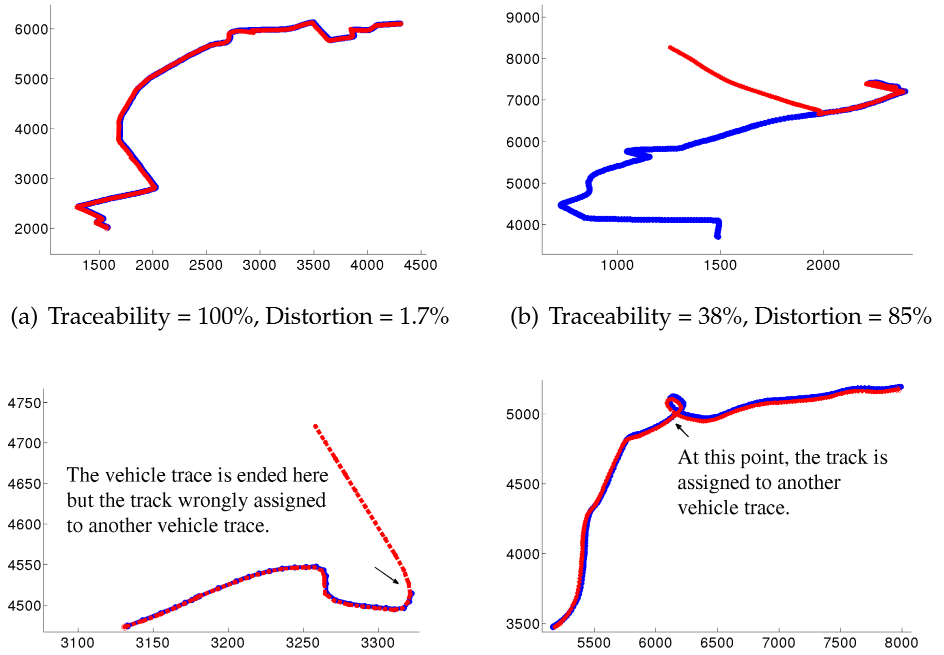
equals zero. This assignment problem is solved using an auction algorithm considering tracks as the bidders, vehicle traces as the items and $l(v,t)$ as the bidding price. After the optimal assignment is obtained, the traceability of the whole scenario is calculated by counting the percentage of significantly tracked vehicles. Thus, the traceability metric Π is defined as:

$$\Pi = \frac{1}{N}\sum_{v\in V}\lambda_v \times 100, \quad \lambda_v = \begin{cases} 1 & \frac{\tau_v}{L(v)} \geq 0.90 \\ 0 & otherwise \end{cases} \tag{4.5}$$

where $L(v)$ is the lifetime of v and N is the total number of traces included in the dataset. This metric allows few confusions around the endpoints of a vehicle trace (10% of the trace lifetime) since inaccuracies in endpoints can be smoothed by a clustering technique in a re-identification process, as shown in [66]. According to this definition, the privacy of the driver is considered breached if the adversary is able to continuously track 90% of her trace. Also, this metric reflects the probability of being tracked by calculating the ratio of tracked vehicles rather than how long a tracker can estimate from the actual trace as in [45, 143].

When the number of traces and tracks are huge, allocating a single assignment matrix for all of them is significantly memory intensive process. To overcome this issue, the traces are divided into time windows of 15 min each so that the traces appear in a window and their corresponding tracks are processed together. The tracks assigned to traces in a time window will never be processed in the subsequent windows. This workaround may lead to a non-optimal solution because it gives a higher priority to former traces for track assignment. However, we compared the assignments obtained from this workaround and those obtained from the optimal method in several tests and we noticed they are almost similar.

There is a shortcoming in measuring privacy using traceability only. The traceability does not consider how distorted the reconstructed tracks if compared to the original traces. In most cases, high traceability indicates low distortion and vice versa because, at the end, tracks are reconstructed from precise and frequent spatiotemporal samples exchanged for safety applications. However, it is not necessarily the case. Figure 4.2 demonstrates four different traces and their assigned tracks showing the traceability and distortion metrics. Figure 4.2(a) presents a case where the entire vehicle trace is reconstructed into a single track, which is never assigned to another vehicle trace, resulting in perfect traceability and very low distortion. Figure 4.2(b) illustrates the case when the assigned track reconstructs only a partial segment of the vehicle trace resulting in low traceability and high distortion. These two examples show the apparent inverse proportionality between traceability and distortion. How-



The vehicle trace is ended here but the track wrongly assigned to another vehicle trace.

At this point, the track is assigned to another vehicle trace.

(a) Traceability = 100%, Distortion = 1.7%

(b) Traceability = 38%, Distortion = 85%

(c) Traceability = 100%, Distortion = 45%

(d) Traceability = 51%, Distortion = 1.8%

Figure 4.2: Traceability and distortion metrics comparison. Each figure illustrates a single vehicle trace drawn in blue and its assigned track drawn in red.

ever, the traceability metric sometimes does not indicate the actual distortion. For example, it can happen that the assigned track is longer than the original trace because the adversary is confused at the end of the vehicle trace and it further associated this track to another vehicle trace, as shown in Figure 4.2(c). In this example, the vehicle trace starts at the bottom left and assigned to the track till the end of the trace lifetime. At this point, the adversary assigned this track further to another vehicle trace (not shown in the figure). The track assignment process assigned this track to the first trace, and thus the traceability metric assumes a perfect tracking because the entire vehicle trace is assigned to a single track. However, the reconstructed track is largely different from the original trace which preserves some privacy and must be reflected in the metric. Another example is when the track is assigned to a partial vehicle trace and then assigned to another near vehicle trace, as shown in Figure 4.2(d). In this example, the trace spatially appears similar to the assigned track because the second vehicle trace follows the same routes of the first trace. Since the adversary is confused in the middle of the trace, the traceability metric shows a partial tracking. However, the privacy is breached indeed since the track spatially reconstructs the whole trace. Therefore, for a better privacy measurement, the distortion of the assigned track should be included in the metric.

The distortion-based metric is measured by calculating how different the assigned track from the original vehicle trace. The tracks are first assigned to vehicle traces so that the total tracking periods are maximized for the whole scenario, as defined in Equation 4.4. Then, the ratio of the distorted segments to the total trace length is calculated to indicate the distortion ratio. Formally, let the track t consist of spatiotemporal samples $t_p, t_{p+1}, ..., t_m$ and it is assigned to the vehicle trace v which consists of spatiotemporal samples $v_q, v_{q+1}, ..., v_n$ (i.e., $t \sim v$) where it is not necessary that $p = q$ or $m = n$. We define the distortion of sample pairs $\delta(v_i, t_i)$ at a time step $i, \forall i, max(p, q) \leq i \leq min(m, n)$ as follows:

$$\delta(v_i, t_i) = \begin{cases} 1 & Ed(v_i, t_i) > \varepsilon \quad or \quad \nexists\, t_i \\ 0 & otherwise \end{cases} \tag{4.6}$$

where $Ed(v_i, t_i)$ is the euclidean distance between v_i and t_i and ε is a distortion threshold. According to this definition, $\delta(v_i, t_i)$ qualifies t_i as distorted if it is farther from v_i by at least ε or the adversary cannot reconstruct the sample v_i (i.e., $\nexists\, t_i$). The distortion threshold ε should be sufficiently large in order to take into account possible distance errors between v_i and t_i. For example, let a track t be assigned to a trace v_a until a time step k and then t is further assigned to another trace v_b, as in Figure 4.2(d). It is likely that v_b lags in time from v_a which leads to a spatial distance between corresponding samples of v_a and v_b at the same time step. These time lag and spatial distance are reflected in the track samples since they are reconstructed from v_b rather than v_a starting from the time step k. We assume a time lag of 5 s or a spatial distance of 75 m is allowed, assuming an average speed of 15 m/s.

The length of the distorted paired segments of t and v is calculated by taking the longest distorted segment from the reconstructed track or the original trace, as follows:

$$\Delta_p = \max\{ \sum_{i=max(p,q)}^{min(m,n)-1} Ed(v_{i+1}, v_i) \cdot \delta(v_i, t_i), \sum_{j=max(p,q)}^{min(m,n)-1} Ed(t_{j+1}, t_j) \cdot \delta(v_j, t_j) \} \tag{4.7}$$

Since the track and the original trace may start and end at different times, a penalty should be added to take these unmatched segments into account. Thus,

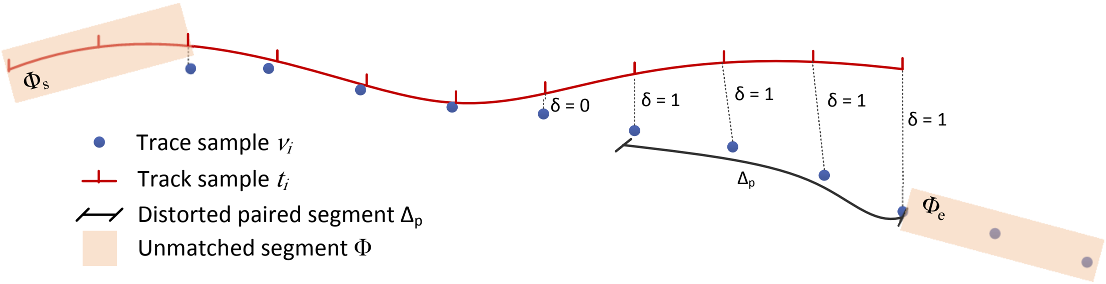


Figure 4.3: Components of the distortion metric

ϕ_s and ϕ_e are defined to count this distortion as follows:

$$\phi_s = \begin{cases} \sum_{i=q}^{p-1} Ed(v_{i+1}, v_i) & p > q \\ \sum_{i=p}^{q-1} Ed(t_{i+1}, t_i) & p < q \\ 0 & otherwise \end{cases}, \phi_e = \begin{cases} \sum_{i=m}^{n-1} Ed(v_{i+1}, v_i) & m < n \\ \sum_{i=n}^{m-1} Ed(t_{i+1}, t_i) & m > n \\ 0 & otherwise \end{cases} \tag{4.8}$$

Figure 4.3 illustrates an example for calculating the distortion for paired and unmatched segments. In this example, the track starts before the beginning of the vehicle trace and ends before the trace end. From their paired samples, there are four distorted samples because their inter-distances are larger than ε. The unmatched segments from the trace and track are highlighted by light orange rectangles.

Given these components, the distortion of the vehicle trace v can be calculated as the ratio of the total length of the distorted segments to the length of the original trace or the length of the reconstructed track, whichever is longer, as follows:

$$D_v = \frac{\Delta_p + \phi_s + \phi_e}{\max\{\sum_{i=q}^{n-1} Ed(v_{i+1}, v_i), \sum_{j=p}^{m-1} Ed(t_{j+1}, t_j)\}} \tag{4.9}$$

The distortion D of the whole scenario can be expressed as the percentage of vehicle traces that their distortion exceeds a specific ratio which guarantees preserving the driver's location privacy (e.g., $D_v > 0.25$). Formally, D can be defined as follows:

$$D = \frac{1}{N} \sum_{v \in V} \alpha_v \times 100, \quad \alpha_v = \begin{cases} 1 & D_v > 0.25 \quad or \quad t \nsim v \quad \forall t \in T \\ 0 & otherwise \end{cases} \tag{4.10}$$

Here, the trace is considered distorted if its D_v is more than 0.25 or there is no track assigned to this trace. We assume that traces distorted by this ratio are not beneficial in posing further privacy attacks. Since the distortion is calculated based on a track that continuously reconstructs the vehicle trace, the distorted segment will be at the trace endpoints. This means that the source and/or destination of the distorted traces cannot be reconstructed which makes re-identification very difficult. Lower distortion ratios may be sufficient as well, but we chose a sufficiently large ratio to ensure a true privacy preserving level.

Furthermore, some vehicles never change their pseudonyms during their lifetime which leads to perfect tracking by matching the same pseudonym. Thus, we additionally measure the *normalized distortion* D_n by excluding these traces. This normalized metric considers the effectiveness of the privacy scheme when a vehicle changes its pseudonym at least once and is defined as:

$$D_n = \frac{1}{N} \sum_{v \in V} \alpha_v^{norm} \times 100, \quad \alpha_v^{norm} = \begin{cases} 1 & \alpha_v = 1 \wedge psd_v(q) \neq psd_v(n) \\ 0 & otherwise \end{cases} \tag{4.11}$$

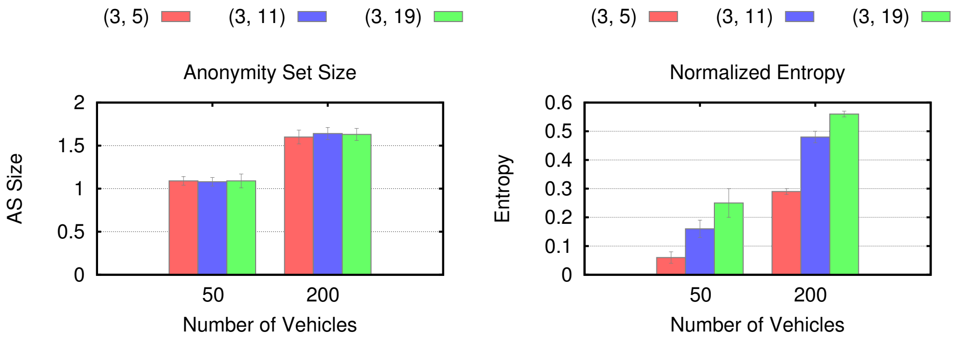
where $psd_v(q)$ and $psd_v(n)$ are the pseudonyms of the trace v at the first and last time steps of its lifetime, respectively.

Based on the metric definitions in Equations 4.10 and 4.11, the distortion is calculated as a ratio of the distorted segment to the total trace length rather than a distance error which provides a unified scale for privacy measurement. Also, the original traces, used as a ground truth, are the actual vehicle traces obtained from the dataset without any noise or silence periods. However, they are trimmed by the time period in which they appear in the dataset obtained from the privacy scheme. Moreover, this metric considers traceability implicitly since the track-to-trace assignment is obtained by maximizing the tracking period for the whole vehicle traces.

4.4 Metrics Comparison

According to the explanation given in the previous section, the distortion seems to be the most representative metric for location privacy. In this section, we provide an experimental comparison among the presented metrics to verify this finding. The experiment consists of applying a simple privacy scheme with three parameter sets, which it is known that they result in low, intermediate and high privacy levels, respectively. We used STRAW vehicle traces presented in Section 1.6.2 in both low and high density scenarios (i.e., 50 and 200 vehicles). A good privacy metric should show reasonable variation among different parameter sets and different densities. We chose the random silent period (RSP) privacy scheme which keeps the pseudonym for a fixed preset time (120 s) and then changes it and keeps silent for a random time period.

We selected random silent periods of (3, 5) s, (3, 11) s and (3, 19) s to achieve low, intermediate and high privacy levels, respectively. We applied the RSP with each parameter set on the traces dataset of each density 10 times. Then, we used the vehicle tracker explained in Section 3.4.7 to track pseudonymous beacons generated by the RSP.



(a)

(b)

Figure 4.4: The AS size and entropy metrics comparison in STRAW vehicle traces.

The traceability and distortion metrics are calculated as defined in Equations 4.5 and 4.10, respectively. For the anonymity set (AS) size, we calculate the maximum AS size encountered by each vehicle and then taking the average over all vehicles. The maximum AS size of a subject vehicle is obtained by finding the maximum number of nearby vehicles, including itself, that changed their pseudonyms simultaneously over each pseudonym changed by this subject vehicle. Two vehicles are considered nearby if they are located within a distance of 100 m. For the entropy, we calculate the maximum normalized entropy $\mathcal{H}_n$, defined in Equation 4.2, of the pseudonym changes made by a vehicle and then take the average over all vehicles.

Figures 4.4 and 4.5 show the results of each metric with the three silent periods in low and high density scenarios. In Figure 4.4(a), the AS size is almost the same in all silent periods with a slight difference between low and high densities. This highlights the inability of the AS size of discriminating the capabilities of different privacy schemes. The normalized entropy overcomes this problem and shows consistent variation among different silent periods, as illustrated in Figure 4.4(b). However, the entropy values are misleading because they do not reflect the true privacy level in different scenarios. For example, the normalized entropy of the RSP (3, 5) in the dense traffic is higher than the RSP (3, 19) in the sparse traffic. This is true regarding the adversary uncertainty since he will be more uncertain in a dense environment due to, for example, the larger AS size. However, the gained privacy of the RSP (3, 5) in the dense traffic is not that high because most of the vehicle traces ($\geq 90\%$) can be reconstructed effectively, as demonstrated next.

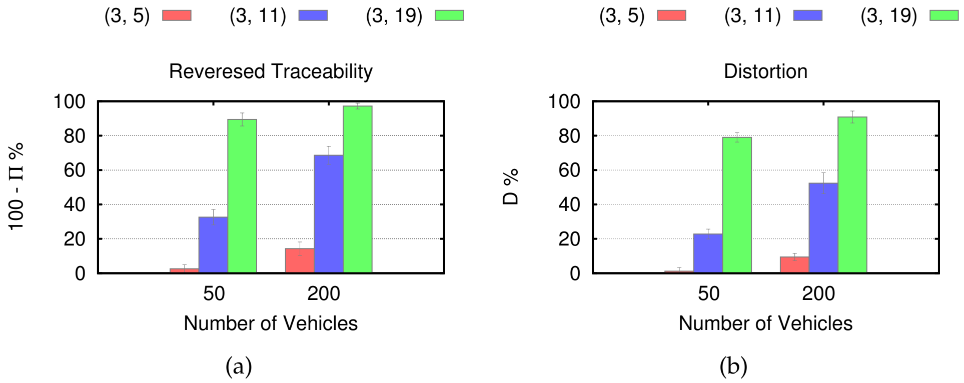


Figure 4.5: The traceability and distortion metrics comparison in STRAW vehicle traces.

In Figure 4.5(a), we show the reversed traceability (i.e., $100 - \Pi$) instead of the traceability metric to reflect the privacy level and be consistently comparable with other metrics. It shows a significantly different variation from that given by the entropy metric. In contrast to the entropy, it demonstrates a low privacy level in the dense traffic when using a short silent period of (3, 5) s. Also, it shows that privacy can be effectively preserved in a sparse traffic when using a relatively long silent period of (3, 19) s. This difference in the variation distribution of the reversed traceability comes from the fact that it measures the effectiveness of reconstructing complete vehicle traces rather than the adversary uncertainty. Last but not least, the distortion metric produces similar variations as the reversed traceability, but it reduces the percentage values indicating lower privacy. This reduction comes from the fact that the distortion metric filters out the cases when vehicles are completely tracked but their reconstructed tracks are still different from the original vehicle traces, as illustrated in Figure 4.2(c).

Furthermore, we repeat the same experiment on the realistic traces described in Section 1.6.3. The obtained results are similar to those presented with the STRAW traces. The AS size is almost the same for all silent periods while the normalized entropy shows consistent variation. The reversed traceability and distortion metrics show a lower privacy level than that achieved in STRAW traces. This may happen because the length of STRAW traces ($\simeq$ 15 min each)

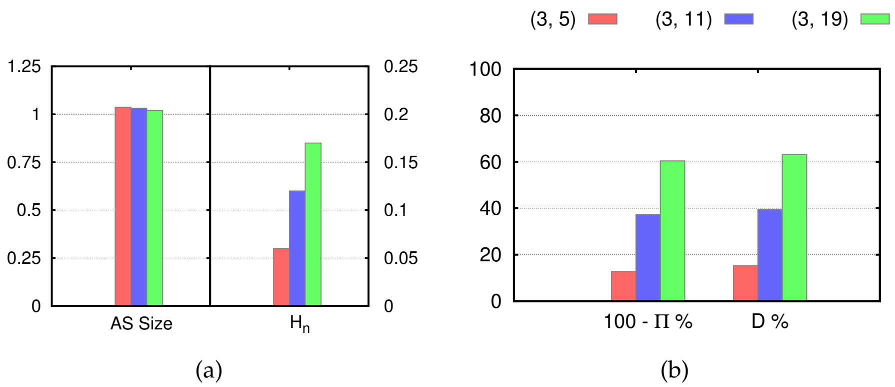


Figure 4.6: Metrics comparison in the realistic vehicle traces.

is longer than the average length of realistic traces ($\simeq$ 5 min). Longer traces result in more pseudonym changes which make them more difficult to be tracked or not distorted which is reflected in the higher reversed traceability and distortion levels for STRAW traces. These results confirm experimentally the suitability of the distortion metric over the other presented metrics to measure location privacy.

4.5 Summary

In this chapter, the location privacy metrics are reviewed in detail and experimentally evaluated. A privacy metric that is based on traceability and distortion is proposed and formally defined. Experiments on both STRAW and realistic vehicle traces showed two main conclusions. First, the anonymity set size and entropy are not suitable location privacy metrics because they do not provide a reasonable protection variation among different privacy schemes in different scenarios. Second, the proposed distortion metric effectively measured the protection level of different privacy schemes on an unified scale. In chapters 6 and 7, the proposed distortion metric will be used to measure the location privacy level of the presented and proposed privacy schemes.

5 Measuring Quality of Service of Safety Applications

5.1 Introduction

Location privacy cannot be preserved with no cost. Privacy mechanisms modify the exchanged information whether by elimination or obfuscation to protect privacy. These modifications affect the *quality of service (QoS)* of the applications. The more constrained the application, the more affected by privacy schemes. For example, safety applications require information about vehicle states frequently, precisely and with lowest latency. Infotainment applications have less restricted constraints. It is important to measure the impact of a privacy scheme on the QoS of applications to ensure they will operate effectively given the information modified by the privacy scheme. We consider safety applications in our QoS analysis because privacy schemes modify beacon messages on which safety applications depend. Besides, safety applications have the most restricted constraints regarding information accuracy, frequency and latency. If a privacy scheme does not hinder the QoS of safety applications, it will not do for other applications as well.

In VANET, the QoS is measured from different perspectives. Most research works measured it in terms of communication parameters, such as packet loss and routing efficiency, or in terms of errors in the received information. Only few researchers who measured the QoS as the expected deficiency of the application operations.

In this chapter, we propose a QoS measurement approach for VANET safety applications given beacon information modified by a privacy scheme. The main concept of this proposal is that a vehicle tracks the movement history of its nearby vehicles to enhance and complement its view on the surrounding traffic. This *in-vehicle tracker* filters errors of the measurements received from other vehicles and estimates their states if their beacon messages are missed. Thus, the QoS should be evaluated considering this enhanced information rather than the received information. Additionally, we assume that an appropriate QoS metric should reflect the deficiency in the application performance rather than absolute distance errors or time delays. The issue in measuring the QoS as a distance error or a time delay is that it does not explain the actual robustness of the application against information inaccuracy. For example, a QoS of one meter error in position does not indicate that the safety application will produce an accurate collision alert because it depends on the application requirements and how they are calculated. Therefore, we propose formulating the application requirements and using Monte Carlo numerical analysis to estimate the QoS given the information enhanced by the in-vehicle tracker. Since we select specific applications for the QoS analysis, we consider cooperative collision warning (CCW) applications as representative for safety applications. They require the most precise (i.e., < 1 m error) and the most frequent (i.e., up to 1 Hz) information about vehicle states [38, 122]. The CCW applications have three distinct types of warnings: forward collision, lane change and road intersection. We present analysis for the first two types; forward collision warning (FCW) and lane change warning (LCW) applications.

5.2 Related Work

Some researchers evaluated the impact of location privacy schemes on the QoS of applications. However, the QoS metric differs from a study to another. In general, the existing QoS metrics can be divided into three categories based on the measured aspect whether the communication quality, data quality (position error) or application requirements. For communication quality aspect, Schoch *et al.* [119] analyzed the impact of pseudonym changes on the performance of geographic routing. Their results confirm serious performance degradation in case of less-density traffic and frequent pseudonym changes ($< 30s$). They suggested introducing a callback mechanism which informs the routing about failed transmissions to cope better with pseudonym changes. Huang *et al.* [72] measured the QoS in terms of the maximum gap within communication and bit rate of information. They used silent period to provide unlinkability for a pseudonym change. Their QoS metric is the silent ratio which is the ratio of silent time to the total time of pseudonym lifetime and silent time. Calandriello et al. [30] measured the impact of pseudonym change in terms of the reception timing of the new pseudonym in several distances and relative speeds.

For data quality metrics, Hoh *et al.* [67] presented a QoS metric for traffic monitoring application characterized as the error applied to each individual location sample. For the metrics based on the application requirements, Hoh *et al.* [69] measured the data quality through the relative weighted road coverage. They considered a road segment covered if a data sample with 100 m accuracy is available. They used several analysis studies for traffic monitoring applications to identify the requirements and constraints for accurate performance. Papadimitratos *et al.* [102] studied the impact of different VANET security and privacy schemes on an emergency braking alarm application. They simulated a dense platoon of vehicles moving with relatively high speed and counted the occurrences of vehicle collisions upon an emergency braking of the leading vehicle. Lefevre et al. [81] analyzed the influence of the duration of the silent period on the effectiveness of intersection collision avoidance (ICA) systems based on VANET. They proposed an ICA system and evaluated a silent period scheme in terms of missed and avoided collisions. They claim that the ICA system can function well with silent periods less than two seconds.

5.3 Proposed Measurement Approach

In this section, we explain the proposed QoS measurement approach. We first describe the measurement methodology and then apply it on two safety applications which are FCW and LCW applications.

5.3.1 QoS Measurement Concept

The main concept of the proposed QoS measurement approach is to formulate the probability of estimating safety application requirements in terms of the vehicle states. Examples of these requirements are correctly identifying the lane of the vehicle and calculating the time-to-collision with a leading vehicle. Monte Carlo numerical analysis is used to calculate these probabilities given the vehicle states which may be obfuscated or eliminated after applying the privacy scheme. Once the probability of each requirement is estimated, all these probabilities are combined to express the QoS metric. The advantage of measuring the QoS in this way is the ability to reflect the realistic performance of each application by considering its requirements with no need to implement it. Other generalized QoS metrics, such as the mean location error [67], are not sufficient because they do not correlate the inaccuracy of the information to the actual operations of the application.

This QoS measurement method is inspired by the approach presented by Shladover and Tan [122] to determine the probability of providing useful CCW warnings as a function of the position and speed accuracy. We apply the same concept with similar assumptions which are as follows:

- The position and velocity obtained from vehicle sensors are erroneous and their errors follow Gaussian distribution.
- To simplify the formulation of the requirements, it is assumed that vehicles are driving on straight roads, centered in their lanes and have constant speed without changing their lane.

- Communication and computation delays are ignored.

These assumptions are considered to simplify the Monte Carlo equations without loss of generality. The second assumption is considered only during instantaneous Monte Carlo calculations. If this assumption were to be removed, the equations would become complex because it would be necessary to consider the vehicle heading, position and velocity in both lateral and longitudinal coordinates [1]. Our novel contribution in this part is the method for obtaining error samples from privacy schemes, as explained next.

To produce stable estimations, Monte Carlo analysis requires a large amount of samples drawn from the random distribution of the measurement errors. As position and velocity measurements are necessarily erroneous and they are sometimes perturbed or eliminated to increase privacy, generating such samples should be performed carefully to reflect the correct representation of the data. Initially, we add a basic noise to positions and speeds specified in the vehicle traces dataset. The basic position noise is drawn from a Gaussian distribution with a standard deviation of 0.5 m. The basic speed noise is assumed to have a Gaussian distribution, and its standard deviation equals 2% of the actual speed. These small errors are recommended in [122], as they lead to a QoS of approximately 95% in CCW applications. Also, they are already realized in systems such as [120] by incorporating information received from a DGPS receiver along with common vehicle sensors.

To estimate the error distribution originating from a privacy scheme, there are two options. First, the safety application depends on instantaneous measurements from other vehicles without keeping track of their movement history. In this case, the quality of service will be directly affected by the amount of the added noise. In addition, the application will not detect the existence of vehicles when their beacons are missed due to a silence period or a communication problem. The second case, which is the one assumed here, the safety application tracks the surrounding vehicles continuously aiming to enhance their measurements and also estimate their state when beacons are missed. In this case, the safety application works like a tracker to track and filter measurements received from other vehicles. Therefore, when evaluating the QoS of a safety application after applying a privacy scheme, we obtain error samples from the vehicle tracker.

The error samples of a privacy scheme are generated as follows and as shown in Figure 5.1. First, the vehicle traces are modified according to the privacy scheme (pseudonyms are changed, beacons are eliminated during silence) to

[1] The lateral and longitudinal coordinates are perpendicular and parallel to the road direction, respectively. In the rest of this chapter, the longitudinal coordinate is referred by x while the lateral coordinate is referred by y.

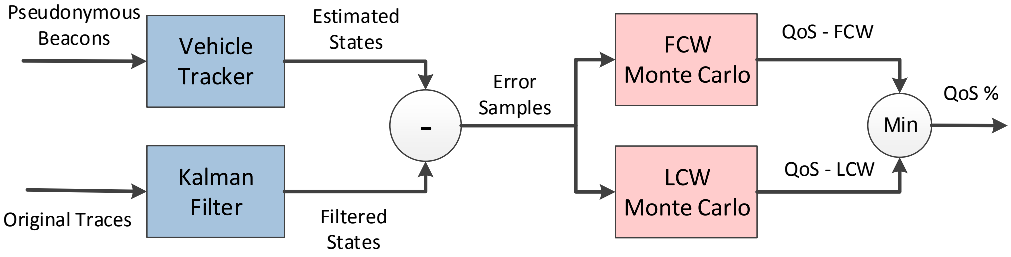


Figure 5.1: Block diagram of the QoS metric

generate pseudonymous beacons. Next, the vehicle tracker operates on theses pseudonymous beacons and tries to reconstruct the original traces. The position and speed errors between the reconstructed tracks and the actual traces are calculated for all vehicles and time steps. These error samples are collected and used directly in the Monte Carlo analysis rather than fitting their error distribution. Thus, this method is generally applicable to any privacy scheme, since the error distribution will differ from a privacy scheme to another. However, the error estimation is customized to our tracker and thus, different trackers and state estimation techniques (e.g., Particle filter) may result in different QoS evaluations. The number of error samples extracted from a single run of the sparsest vehicle traces equals to one hundred thousand samples, which is a sufficient number to obtain stable Monte Carlo results. In very large datasets, we obtain only half a million of error samples distributed over time steps.

The actual traces used in calculating error samples are slightly different from the original traces in datasets. Generally, the Kalman filter modifies the position and speed from those recorded in the traces dataset as a kind of enhancement even if no noise or privacy scheme is applied. These enhancements will contribute to the extracted error samples if the original traces are used. Thus, we calculate the error samples by taking the *filtered traces* as the ground truth. These filtered traces are obtained by applying the Kalman filter on each vehicle trace individually and taking the position and speed of the estimated state every time step. Thus, the error samples are guaranteed to originate from changes made by the privacy scheme only, not from changes made by the Kalman filter. Moreover, the error samples are measured in the scenario global coordinate, but, according to our assumptions, they are needed to be in the vehicle coordinates (i.e., lateral and longitudinal), as explained in the next sections. Therefore, the coordinate system of the measurements are rotated by the instantaneous vehicle heading, assuming it drives in the same direction as the road, before calculating the error. The error sample Δ is formally calculated as follows:

$$\Delta = \begin{bmatrix} \delta x \\ \delta \dot{x} \\ \delta y \\ \delta \dot{y} \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \cdot (\hat{\mathbf{x}}_p - \hat{\mathbf{x}}_f), \quad \mathbf{R} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{5.1}$$

where θ is the vehicle heading, $\hat{\mathbf{x}}_p$ is the estimated vehicle state by the inside tracker and $\hat{\mathbf{x}}_f$ is the filtered state. Both $\hat{\mathbf{x}}_p$ and $\hat{\mathbf{x}}_f$ consist of position and velocity in xy global coordinate. The block diagram of the QoS metric calculation is illustrated in Figure 5.1. Next, we will show how these error samples are used to estimate the QoS of the FCW and LCW applications.

5.3.2 Forward Collision Warning Application

The Forward Collision Warning (FCW) application aims to provide the driver of the subject vehicle (SV) a sufficiently early alert that a possible collision with another vehicle (OV) in the same lane is likely, as shown in Figure 5.2. The SV is the vehicle equipped to give the warning, and the OV is any other vehicle. To achieve this functionality, the application must be able to (1) identify the correct lane of OVs and (2) estimate the time to collision (TTC) within a small tolerance. To satisfy the first requirement, accurate lateral positions of the SV and OVs must be known. To satisfy the second requirement, knowledge of the longitudinal positions and speeds of the SV and the next OV in the same lane is necessary. In our analysis, we assume that the errors of the SV measurements are just the basic error in position and speed as the SV obtains these values through its own sensors, rather than the VANET communication.

For the first application requirement, the SV must correctly identify that OV1 is in its own path (i.e., high sensitivity) while OV2 is not (i.e., high specificity), as shown in Figure 5.2. The criteria for identifying an OV as in path is that its lateral position is within ± 1.8 m of the lateral position of the SV, assuming a 3.6 m lane width. Otherwise, it should be identified as not in path. In our analysis, we set the true lateral position of the SV as same as the lateral position of OV1, while the true position of the OV2 is located in the center of the next lane. Thus, the measured lateral positions of SV, OV1 and OV2 are obtained by adding the errors to their true positions as follows:

$$\begin{aligned} y_{SV} &= 1.8 + \mathcal{N}(0, 0.5) \\ y_{OV1} &= 1.8 + \delta y \\ y_{OV2} &= 5.4 + \delta y \end{aligned} \tag{5.2}$$

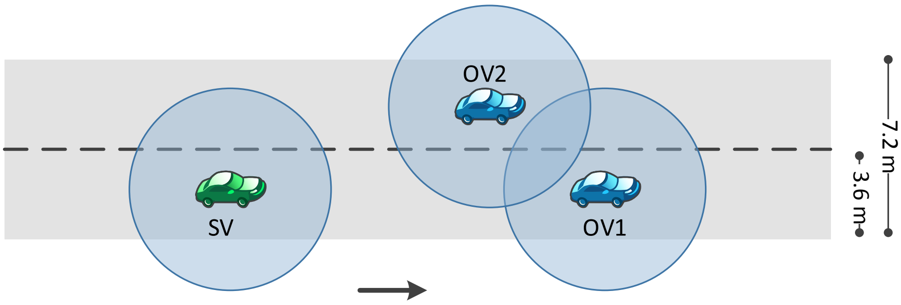


Figure 5.2: Forward collision warning scenario

Therefore, the true and false positive probabilities for correctly identifying lanes of the OVs can be calculated by:

$$P_{true+} = P(|y_{OV1} - y_{SV}| \leq 1.8) \tag{5.3}$$

$$P_{false+} = P(|y_{OV2} - y_{SV}| \leq 1.8) \tag{5.4}$$

For the second requirement, we assume that the SV is approaching the OV1 at speed differences Δs of 5 m/s and 15 m/s. The assumed true TTC is set to three seconds as an example; thus, the true position of OV1 is generated to be three seconds distance from the true position of SV and is calculated based on the evaluated speed difference as follows:

$$\begin{aligned} x_{SV} &= \mathcal{N}(0, 0.5) \\ x_{OV1} &= 3 \cdot \Delta s + \delta x \\ \dot{x}_{SV} &= \hat{x}_{OV1} + \Delta s + \mathcal{N}(0, 0.02 \cdot (\hat{x}_{OV1} + \Delta s)) \\ \dot{x}_{OV1} &= \hat{x}_{OV1} + \delta \dot{x} \end{aligned} \tag{5.5}$$

where $\hat{x}_{OV1}$ is the filtered longitudinal speed of the OV1. Here, there is no binary classification to calculate false positives; instead, we calculate the probability of calculating TTC within a small tolerance of 500 ms. This 500 ms tolerance is chosen by Shladover and Tan [122] as the maximum tolerance for issuing a useful warning. They also analyzed the implication of a desirable tolerance of 200 ms but they found that it requires a positioning accuracy of 20 cm to attain this restrict tolerance, wherefore we considered only the maximum tolerance of 500 ms. Therefore, the TTC and the probability of correctly estimating it within 500 ms can be calculated by:

$$TTC = \frac{x_{OV1} - x_{SV}}{\dot{x}_{SV} - \dot{x}_{OV1}} \tag{5.6}$$

$$P_{TTC} = P(|TTC - 3| \leq 0.5) \tag{5.7}$$

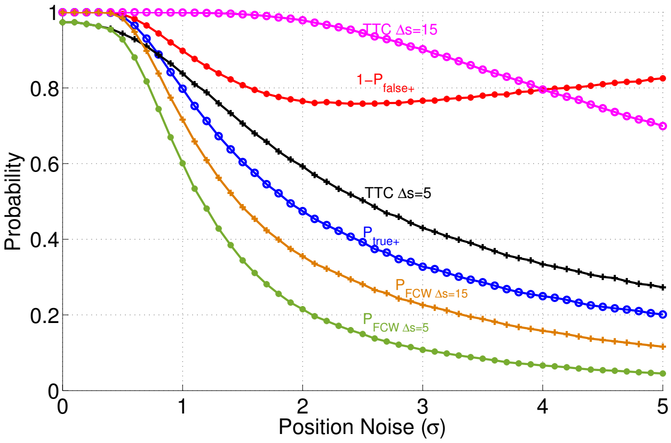


Figure 5.3: Probability of correctly estimating the FCW requirements using positions of Gaussian noise

In this equation, we determine how frequently the difference between the calculated TTC and the true TTC (i.e., 3 s) is less than the tolerance threshold of 0.5 s. Finally, the probability of the FCW application (P_{FCW}) can be obtained by multiplying all three probabilities together, assuming they are independent, as follows:

$$P_{FCW\Delta s} = P_{true+} \times (1 - P_{false+}) \times P_{TTC\Delta s} \tag{5.8}$$

As a kind of verification with the results shown in [122], we show the behavior of the various probabilities when the position noise is Gaussian in Figure 5.3. The error samples used in this example are not filtered by Kalman filter as described in Section 5.3.1 but they are basically sampled from a Gaussian distribution of the given standard deviation along the x-axis. The obtained results are similar to those presented in [122]. It is worth to note that the position error must be at most 50 cm to achieve P_{FCW} of 0.93 or higher in the FCW application when the speed difference Δs equals to 5 m/s. The governing factor is the P_{TTC} in low noise values ($\sigma < 0.9$ m). Also, it can be observed that estimating TTC in high speed differences is much more accurate than low speed differences with the same position noise. Therefore, the QoS of the FCW application (QoS_{FCW}) is defined as $P_{FCW\Delta s=5}$ multiplied by 100 to obtain a percentage, as follows:

$$QoS_{FCW} = P_{FCW\Delta s=5} \times 100 \tag{5.9}$$

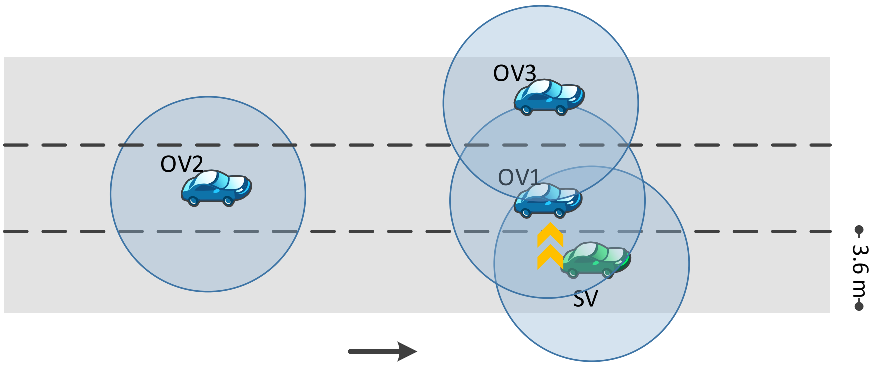


Figure 5.4: Lane change warning scenario

5.3.3 Lane Change Application

There are two main scenarios that lane change warning (LCW) application concerns which are *blind spot* and *overtaking*, as shown in Figure 5.4. In the blind spot scenario, the OV1 moves in the adjacent lane of the SV at approximately the same speed and slightly behind it which poses a threat of collision when the SV changes its lane. Therefore, the LCW application deployed in the SV should give an alert about OV1, but not about OV3 as it is located in the third lane and does not threat the SV. In the overtaking scenario, the approaching OV2 comes from the rear with a high closing speed such that it arrives adjacent to the SV in the same time of lane change. If the OV2 is moving in a speed allows it to reach the adjacency of the SV in the time of lane change, then a warning should be issued as it is an overtaking threat. By this illustration, the overtaking scenario is just like that of forward collision warning described before, but the positions of SV and OV are reversed. Thus, we will analyze the blind spot scenario here.

To handle the blind spot scenario, three requirements must be correctly identified by the SV. The first requirement is to identify the lateral position of OV1 in the adjacent lane (i.e., its true center is 3.6 m away from the SV). Additionally, its longitudinal position should be estimated slightly behind the SV, let's say between 1.5 m and 6 m far from the longitudinal position of the SV. Thus, its true longitudinal position is assumed to be in the middle of this range (i.e., 3.75 m from the SV). The second requirement is to recognize the OV3 as not located in the adjacent lane which means its true lateral position is 7.2 m away from the SV. The last requirement is that the speeds of OV1 and SV should be recognized to be similar up to a small margin of 3 m/s as an example. Therefore, the true speeds of SV and OV1 are assumed to be the same. According to these requirements, the measured positions and speeds of SV, OV1 and OV3 are defined as follows:

$$\begin{aligned} y_{SV} &= 1.8 + \mathcal{N}(0, 0.5) \\ x_{SV} &= 3.75 + \mathcal{N}(0, 0.5) \\ \dot{x}_{SV} &= \hat{x}_{SV} + \mathcal{N}(0, 0.02 \cdot \hat{x}_{SV}) \\ y_{OV1} &= 5.4 + \delta y \\ x_{OV1} &= \delta x \\ \dot{x}_{OV1} &= \hat{x}_{OV1} + \delta \dot{x} \\ y_{OV3} &= 9 + \delta y \end{aligned} \qquad (5.10)$$

where $\hat{x}$ is the filtered longitudinal speed and $\hat{x}_{SV} = \hat{x}_{OV1}$. The Monte Carlo equations of each requirement need some further analysis. Assuming 2 m wide SV and OV1, the OV1 must leave enough space for the SV to enter the adjacent lane. This means when the SV changes its lane, the center of the OV1 should be 3 m away from the right edge of the lane. Thus, the warning of a blind spot should be fired if the estimated distance between SV and OV1 less than or equal to 4.8 m. To avoid a false alert about OV3, assume a 3 m wide vehicle moving just along the edge of the third lane. Then, its center is 1.5 m away from the lane boundary. Thus, when the distance between centers of SV and OV3 is more than 6.9 m, the system must not warn. Therefore, the true positive probability is calculated when the OV1 is estimated within a distance less than 6.9 m. The false positive probability is calculated when the OV3 is estimated within a distance less than or equal 4.8 m. Additionally, the longitudinal position of OV1 must be estimated within the blind spot so that it is not easily visible to the SV driver (i.e., 1.5 - 6 m behind the SV). Also, the speeds of SV and OV1 should be estimated to be similar within small tolerance of 3 m/s. These probabilities can be formulated as follows:

$$P_{true+} = P(y_{OV1} - y_{SV} < 6.9) \qquad (5.11)$$

$$P_{false+} = P(y_{OV3} - y_{SV} \leq 4.8) \qquad (5.12)$$

$$P_{long} = P(x_{SV} - x_{OV1} < 6 \wedge x_{SV} - x_{OV1} > 1.5) \qquad (5.13)$$

$$P_s = P(|\dot{x}_{OV1} - \dot{x}_{SV}| \leq 3) \qquad (5.14)$$

The probability of the LCW application (P_{LCW}) can be obtained by multiplying these probabilities together, assuming they are independent as follows:

$$P_{LCW} = P_{true+} \times (1 - P_{false+}) \times P_{long} \times P_s \qquad (5.15)$$

In Figure 5.5, we show the behavior of the various probabilities when the position noise is Gaussian. The error samples used in this example are not filtered by Kalman filter as described in Section 5.3.1. They are basically sampled from a Gaussian distribution of the given standard deviation along the x-axis.

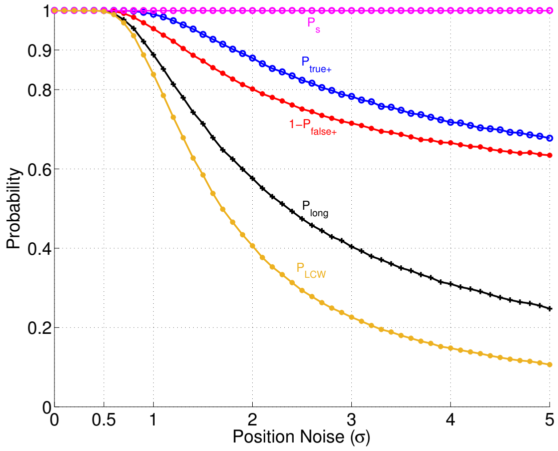


Figure 5.5: Probability of correctly estimating the lane change requirements using positions of Gaussian noise

They are slightly different from those presented in [122] because the authors used unexplained criteria when calculating P_{true+} and P_{false+} on their corresponding figures. According to results in Figure 5.5, the position error must be at most 80 cm to achieve a P_{LCW} of 0.94 or higher. The governing factor is the P_{long} which needs a position accuracy of 90 cm error at most to achieve an accuracy of 0.92 in estimating the longitudinal position. Compared to the requirements of the FCW application shown in Figure 5.3, the LCW application requires slightly relaxed accuracy requirements. Last but not least, the QoS of the LCW application (QoS_{LCW}) is defined as P_{LCW} multiplied by 100 to obtain a percentage, as follows:

$$QoS_{LCW} = P_{LCW} \times 100 \tag{5.16}$$

To measure the impact of a privacy scheme on the QoS of safety applications, both QoS_{FCW} and QoS_{LCW} are calculated, and then the minimum value is taken to express on the final QoS. Formally, the QoS of a privacy scheme is defined as:

$$QoS = min\{QoS_{FCW}, QoS_{LCW}\} \tag{5.17}$$

5.4 Experiment Results

The proposed QoS measurement approach is applied on the FCW and LCW applications and evaluated in two scenarios. The first scenario considers STRAW

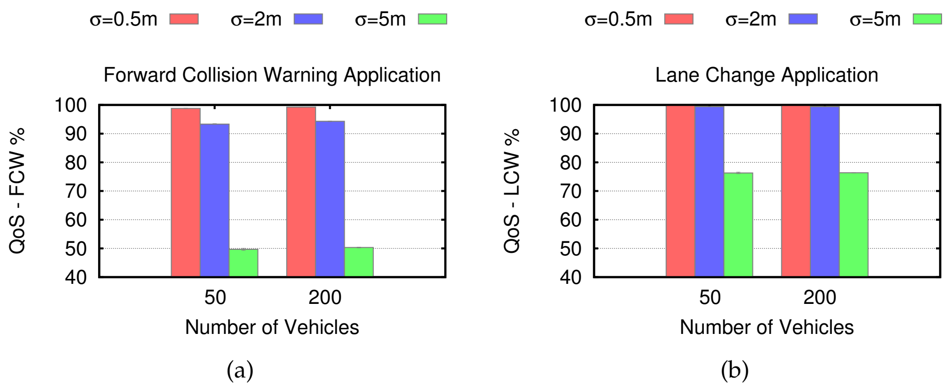
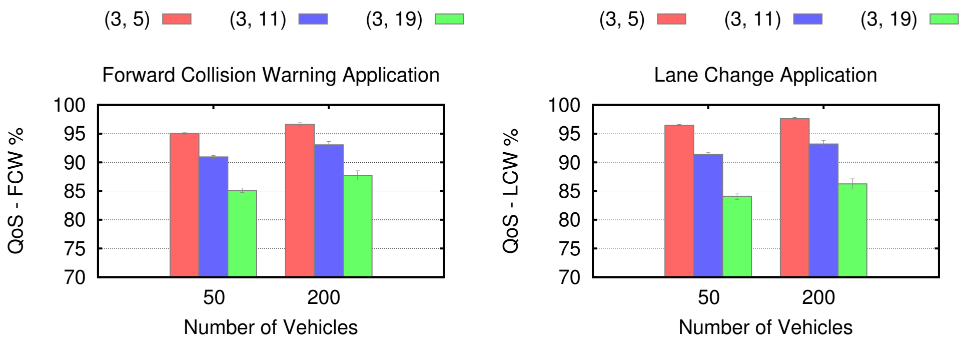


Figure 5.6: The QoS of FCW and LCW applications in noisy STRAW traces

vehicle traces with various Gaussian noises where pseudonyms are periodically changed every two minutes with no silence periods or mix-zones.

In Figure 5.6, it can be noticed that a QoS of 93% or more can be achieved in a FCW application in both sparse and dense traffic when the vehicle position noise is up to 2 m. A better QoS of 99% can be achieved in a LCW application in similar conditions. These QoS values are much higher than those shown in Figures 5.3 and 5.5. This dramatic improvement of QoS comes from the assumption that a vehicle tracks and filters measurements received from nearby vehicles. This in-vehicle tracker filters the position and speed noises and allows both applications to better estimate the states of other vehicles. However, it does not work with large noises (e.g., $\sigma \geq 5$ m) because the tracker is significantly confused and it cannot assign vehicle states to their tracks correctly due to noise.

The second scenario applies the random silent period (RSP) privacy scheme on STRAW vehicle traces. The RSP keeps the vehicle pseudonym for a fixed preset time of two minutes. Then, it changes the pseudonym and keeps silent for a random time chosen from a given period. We selected silent periods of (3, 5) s, (3, 11) s and (3, 19) s to achieve low, intermediate and high privacy levels, respectively. A normally distributed position noise of standard deviation 0.5 m is added before applying the privacy scheme. Figure 5.7 shows that a QoS of 91% or higher can be achieved in both safety applications and both traffic densities if a silent period of (3, 11) s or less is used before a pseudonym change. Moreover, it can be observed that the QoS of the LCW application is slightly higher than that of the FCW application with silent periods up to (3, 11) s. This behavior is reversed with relatively long silence of (3, 19) s. Therefore, in further experiments presented in next chapters, we calculate the QoS of both



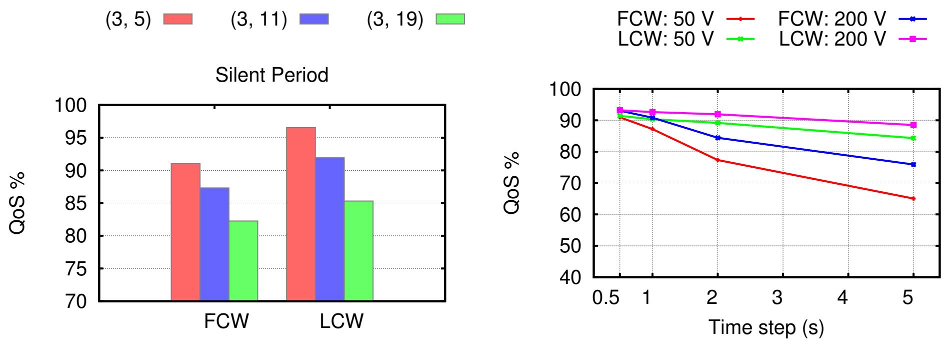
(a)

(b)

Figure 5.7: The QoS of FCW and LCW applications in STRAW traces modified by random silent period privacy scheme and beacon interval = 0.5 s

applications and take the minimum as the QoS of safety application. The finding of this experiment is different from results claimed by Lefevre *et al.* in [81]. They claim that an intersection collision system can function well with silent periods less than two seconds. This difference comes from our proposal that a vehicle tracks and filters measurements received from nearby vehicles. This finding combined with results shown in Figure 4.5 confirms our hypothesis of that it is possible to preserve location privacy without hindering the QoS of safety applications. For example, the RSP of (3, 19) s can achieve a privacy level of 80% in terms of tracking distortion with a loss of about 15% in the QoS of safety applications. Advanced privacy schemes will compromise this trade-off more effectively, as explained in next chapters.

The second scenario is applied on realistic traces to confirm the achieved QoS levels, as shown in Figure 5.8(a). It can be observed that the QoS in both applications are slightly lower than those shown with the STRAW traces. This reduction in QoS occurs because the time step in the realistic traces is 1 s while it is 0.5 s in the STRAW traces. This longer time step prevents the in-vehicle tracker from obtaining the desired accuracy especially for estimating the speed. We verified this finding by testing the STRAW traces with various time steps (i.e., 0.5 - 5 s) and RSP of (3, 11) s, as shown in Figure 5.8(b). It is noticed that the QoS of safety application decreases with longer time steps especially the FCW application. However, the QoS of the LCW application does not decrease significantly. For example, the QoS decreased only up to 4% when a beaconing time of 2 s is used instead of 0.5 s even with using a random silence period of (3, 11) s. This is an interesting finding because decreasing the beaconing rate enhances the network performance.



(a)

(b)

Figure 5.8: (a) The QoS of FCW and LCW applications in realistic traces modified by RSP. (b) The QoS of FCW and LCW applications in STRAW traces with different time steps and silent period of (3, 11) s

5.5 Summary

In this chapter, a measurement approach is proposed to determine the impact of a privacy scheme on the QoS of safety applications. The proposed approach is applied on two applications; forward collision warning (FCW) and lane change warning (LCW) applications. This approach depends mainly on the assumption that vehicles employ a local tracker to track and predict the movement of the nearby vehicles. According to the experiment results, this local tracker enhances the expected QoS of safety applications. For example, in the presence of position noise of 2 m, a vehicle can estimate the requirements of the FCW application by a probability of up to 92% when using this approach, while this probability decreases to only 20% when the vehicle depends directly on the noisy data. Also, using the in-vehicle tracker may relax the requirement of frequent beaconing rate of some applications. Using Monte Carlo analysis, the QoS is measured by calculating the probability of correctly estimating the requirements of the safety application. Finally, the QoS of both applications is evaluated for the random silent period privacy scheme using STRAW and realistic traces. In the following chapters, the QoS of safety applications is considered to be the minimum QoS of the FCW and LCW applications, as defined in the Equation 5.17.

6 Obfuscation Privacy Schemes

6.1 Introduction

As discussed in Section 2.6, location privacy in VANET is usually preserved through changing pseudonyms in an unmonitored area whether a silent period or a mix-zone. Although obfuscation mechanisms are so popular in other domains such as location-based services, they are rarely applied in VANET to avoid degrading the QoS of safety applications [17]. However, by analyzing the actual requirements of safety applications, it can be observed that they will not be entirely dependent on VANET information, but they will also use information that is sensed by the subject vehicle itself. This assumption is valid because VANET will not be penetrated into all vehicles in the initial deployment phase and applications will be designed based on this fact. Also, safety applications cannot guarantee the accuracy of the information received from VANET because of the variation in sensors' accuracy; thus, they have to combine it with self-sensed information. Additionally, safety applications may not continuously require precise updates, except for situations that may concern safety. However, precise and frequent spatiotemporal is required continuously to effectively track vehicles, as discussed in Chapter 3. From this viewpoint, we hypothesize that obfuscation privacy schemes can be applied in VANET if they only add a sporadic noise to position in beacons or change the beacon frequency slightly. These minor modifications may enhance the privacy level without affecting the operations of safety applications significantly.

In this chapter, we propose and evaluate two privacy schemes: (1) position perturbation after a pseudonym change and (2) random beaconing rate. We first discuss related work in Section 6.2. In Section 6.3, the system and adversary models are revisited and important notations are introduced. In Section 6.4, the proposed schemes are presented and evaluated in comparison with random silent period scheme. Last but not least, an advanced adversary is presented and the proposed scheme is reevaluated against this adversary in Section 6.5.

6.2 Related Work

Obfuscation privacy schemes are commonly used to preserve privacy in location based services. One popular privacy mechanism is to degrade the resolution of location information under restriction of the application requirements. For example, spatial cloaking [61] obfuscates the exact location into a region to meet predefined anonymity constraints, such as k-anonymity. Moreover, Hoh and Gruteser [67] perturb user paths so that the total distance error estimated by the adversary is maximized but constrained by a obfuscation radius allowed by the application. This algorithm perturbs traces by adding artificial crossings between near parallel paths to confuse the tracker in an offline central way, where true traces of all users are processed together in a proxy server. Hoh *et al.* [69] propose to hide location samples that lead the adversary to track vehicles for a long period without enough confusion. This algorithm operates iteratively until the optimal location set is reached that confuses tracker. In VANET domain, Wei and Chen [139] perturbed the beacon information (e.g., position, velocity and heading) within a safe radius. This radius is calculated based on the safety conditions with the surrounding vehicles. In our proposed perturbation scheme, a random noise is added to positions for random period after a pseudonym change to provide unlinkability between beacons of new and old pseudonyms.

Adaptive beaconing rate is often employed in VANET but to enhance the network performance not for privacy preserving purposes [118]. Fukui *et al.* [56] proposed to send beacons periodically based on a constant distance a vehicle has to travel. Also, the beacon rate should be reduced when a high node density or a higher packet error rate are detected. Khorakhun *et al.* [79] adapt the beacon rate depending on the current channel load, evaluated through the channel busy ratio. To adapt beaconing rate smoothly, beacon rates calculated by individual vehicles are exchanged among vehicles. Then, the average rate is calculated and applied by each vehicle. Rezaei *et al.* [113] adapt the beaconing rate depending on differences from position predictions. They assume that all vehicles run an extended Kalman filter locally for each nearby vehicle. It continuously estimates the current position based on the received beacons. The vehicle sends a beacon only when there is a difference between its actual position and the remote estimator.

6.3 System and Adversary Models

We assume a system model similar to that defined in Section 1.5.1. Beacon messages are broadcast with a beaconing rate $1/t_b$. Pseudonyms are changed every

Table 6.1: Table of Notations

Notation	Description
t_b	Default beaconing time between every two consecutive beacons required by the application (default beaconing rate = $1/t_b$)
t_p	Pseudonym lifetime
$t_{c(i)}$	Time series at which pseudonyms are changed, $i = 1, 2, 3, ...$
t_s	Tracker sampling time
t_r	Random multiplier used to broadcast beacons on a random basis ($t_r \in \mathbb{N}$)
d_n	Random multiplier where position is noised during $d_n \cdot t_b, d_n \in \mathbb{N}$
d_s	Random multiplier where a vehicle is silent during $d_s \cdot t_b, d_s \in \mathbb{N}$
$\mathcal{N}(\mu, \sigma)$	Gaussian random distribution with mean (μ) and standard deviation (σ)
$U(\alpha, \beta)$	Uniform random distribution, where $a \in \mathbb{N}, \forall a \in U(\alpha, \beta)$. Also noted as (α, β)

fixed time t_p at times $t_{c(i)}, i = 1, 2, 3....$ To prevent synchronization among vehicles, they start with a random pseudonym lifetime which ranges from 1 to t_p.

The assumed adversary model is similar to that defined in Section 1.5.2. Additionally, the adversary may consider all or a selective set of the broadcast beacons by using a *sampling time* t_s (larger than the beaconing time t_b) to skip noise or silence periods. Furthermore, it is assumed that the adversary cannot enhance the broadcast information in a pre-tracking process by matching positions on road maps or by localizing vehicles in the physical layer. It is not expected that these localization methods will produce more accurate information than the broadcast information in beacons. Table 6.1 summarizes the notations used throughout this chapter.

6.4 Proposed Privacy Schemes

In this section, the obfuscation schemes are presented and evaluated. In evaluation, the tracker presented in Section 3.4.6 is employed to reconstruct traces from the beacons modified by a privacy scheme. We employ the sparsest scenario (50 vehicles) of the STRAW vehicle traces described in Section 1.6.2, unless specified otherwise. This low density is selected because the tracker can track vehicles accurately even in the presence of noise, as shown in Section 3.4.6. Thus, it will be a challenge for the privacy scheme to reduce tracking vulnerability in this low density scenario.

6.4.1 Position Perturbation Scheme

The position perturbation scheme adds a large random noise $n \in \mathcal{N}(0, \sigma)$ to a position P_k after a pseudonym change for a relatively short period $d_n \cdot t_b$ where d_n is an integer sampled from a random noise period $U(\alpha, \beta)$. Thus, the modified position $\hat{P}_k$ included in a beacon sent at time k is calculated by:

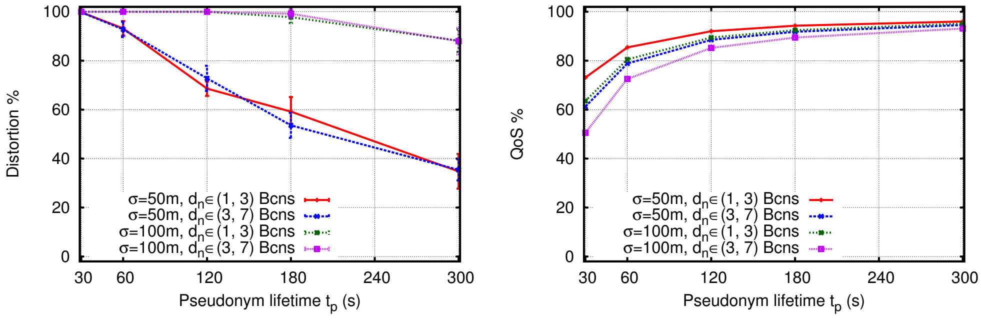
$$\hat{P}_k = \begin{cases} P_k + n & t_{c(i)} < k \leq t_{c(i)} + d_n \cdot t_b \\ P_k & t_{c(i)} + d_n \cdot t_b < k \leq t_{c(i+1)} \end{cases} \tag{6.1}$$

where $t_{c(i)}$ and $t_{c(i+1)}$ are times of two consecutive pseudonym changes. We aim to replace the relatively long silence period commonly used before a pseudonym change with a relatively short noisy period. It is assumed that noisy updates are better from the application perspective than silence. The noisy random period prevents the tracker from skipping the noisy positions when beacons with a new pseudonym are encountered. Intuitively, no noise is added to positions while the same pseudonym is employed, because the tracker will correlate beacons by matching pseudonyms and the added noise will be filtered by the Kalman filter. In real applications, this large noise should be carefully added to be aligned to a realistic location (e.g., aligned to a parallel road) to prevent the tracker from filtering it.

This scheme aims to force the tracker to confuse and assign the beacon of a new pseudonym to another track, which differs from the track assigned to the vehicle trace before the pseudonym change. Note that the tracker cannot fix itself when the noisy period ends, because subsequent beacons are identified by the same new pseudonym and the tracker matches each beacon to the wrong track by pseudonym matching. If the assignment decision of the tracker is modified to be a weighted average of pseudonym matching and the spatiotemporal association, it is advantageous for privacy protection. This weighting causes the tracker to abandon the advantage of simple pseudonym matching and transforms the problem to (partial) anonymous beacons tracking, which is challenging, as discussed in Section 3.4.6.

In Figure 6.1(a), the position perturbation scheme is evaluated with variation of added noises σ of 50 and 100 m and random noise periods of (1, 3) and (3, 7) beacons[1]. The distortion is significantly increased in proportional with the amount of the added noise, while the random noise period has almost no effect. When the added noise is very large of 100 m, the tracker cannot correlate

[1] For simplicity, we will use this notation (α, β) to express $U(\alpha, \beta)$ which should not be misinterpreted as reference numbers. The notation of [a,b] is used for the latter case.



(a) Privacy

(b) QoS

Figure 6.1: Privacy and QoS levels of the position perturbation scheme.

beacons of new pseudonyms because they are far away from their predicted states. For the QoS of safety applications, both the added noise and random noise period affect the QoS, as illustrated in Figure 6.1(b). The QoS decreases with the increase of the added noise and the length of the noise period because the in-vehicle tracker cannot correctly estimate the application requirements using the noised measurements. However, the QoS can be enhanced by increasing the pseudonym lifetime because the ratio of precise to noisy positions is increased which is reflected in a better estimation of the application requirements.

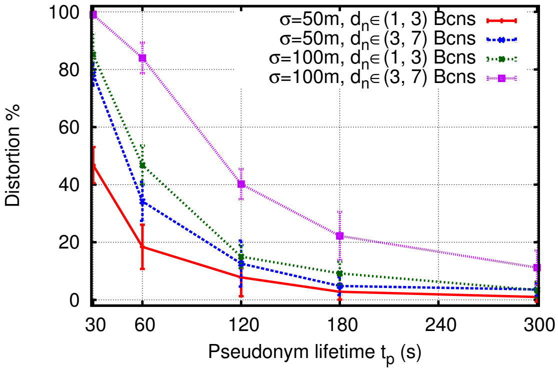


Figure 6.2: Privacy of the position perturbation scheme with a tracking sampling t_s of 7s.

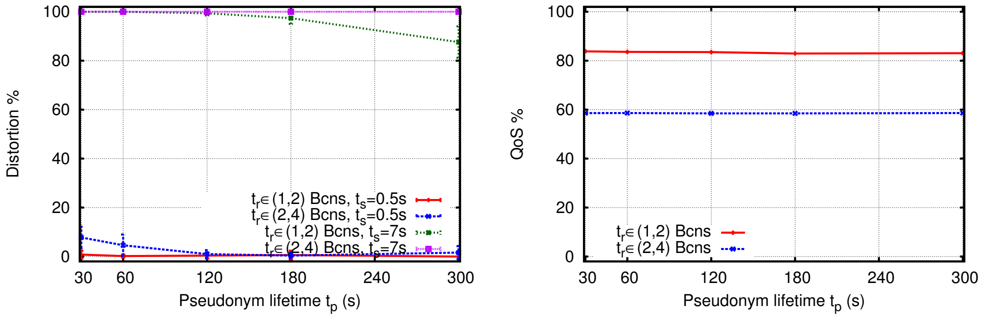
This result is obtained when the tracker considers all beacons sent from vehicles, that is the *sampling time* is the same as the beaconing time (i.e., $t_s = t_b =$

0.5 s). We evaluate the case in which the tracker attempts to avoid the noise period by using a longer sampling time. Figure 6.2 shows the distortion for the same variations of the added noises and noise periods when the tracker uses a sampling time of 7 s. The distortion is reduced significantly especially in longer pseudonym lifetimes ($t_p \geq 120$ s). Note that the QoS is always reduced when $t_s > t_b$ because the in-vehicle tracker cannot estimate the state of surrounding vehicles correctly over this low update rate. These results show the ineffectiveness of the position perturbation scheme when the tracker uses long sampling times.

6.4.2 Random Beaconing Rate

The second privacy scheme is the random beaconing rate. It lets a vehicle broadcast a beacon within a predetermined interval, but selects randomly when the vehicle broadcasts the next beacon (in increments of t_b). Formally, if beacons are basically sent every beaconing time t_b, they will be sent instead every random time $t_r \cdot t_b$ where t_r is an integer sampled from a random beaconing interval $U(\alpha, \beta)$. For example, if the application basically requires sending a beacon every 0.5 s and a random beaconing interval $U(1, 4)$ is allocated, then the vehicle uniformly chooses a random beaconing time $t_r \cdot t_b$ among 0.5, 1, 1.5 or 2 s to send the next beacon. After broadcasting this beacon, it will select another random integer $t_r \in U(1, 4)$ to send the next beacon. The uniform random distribution is used here to ensure that all beacon times have the same probability. This scheme exploits a tracker constraint that requires measurements to be provided on a fixed timing basis for tracked targets. The majority of advanced tracking algorithms can afford few missed measurements of tracked targets. However, if these misses occur regularly and frequently, then the tracks of missed measurements will be mixed with the newly appeared ones at each time step, which reduces the tracking vulnerability. Because the tracker assumes a fixed time step for all vehicles and the scheme continuously changes the beaconing time, the tracker thus must estimate a time step that produces the best tracking results. Because beacons are identified by pseudonyms, the reduction in the tracking vulnerability is likely to be small but beneficial in some cases. One example case involves a beacon of a newly appeared vehicle near a previously encountered vehicle that missed its beacon. In this case, the tracker may mix the beacon of new vehicle with the track of the encountered vehicle because it thinks that the vehicle changed its pseudonym. This situation may generate additional mixes in subsequent time steps, which may reduce the tracking vulnerability.

Figure 6.3(a) shows two random beaconing intervals of (1, 2) and (2, 4) beacons tracked using two tracking samples of 0.5 and 7 s. This scheme does not



(a) Privacy

(b) QoS

Figure 6.3: Privacy and QoS levels of the random beaconing time scheme.

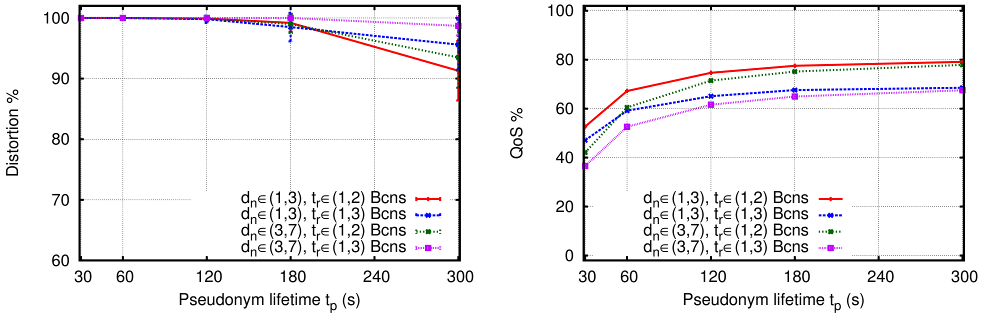
provide any reduction in the distortion when the tracker uses the basic sampling time of 0.5 s. However, we notice that the distortion significantly increases in the other sampling time; exactly the reverse behavior of the position perturbation scheme. Figure 6.3(b) shows the impact of these random beaconing intervals on the QoS. We notice that the impact of the random interval is fixed for all pseudonym lifetimes because this scheme eliminates beacons independently of the pseudonym change. Additionally, the random interval of (2, 4) beacons results in a much lower QoS than the interval of (1, 2) beacons because it significantly eliminates a reasonable number of beacons every time step. These missed beacons prevent the in-vehicle tracker from estimating the application requirements correctly. These results suggest combining both schemes so that the distortion is not reduced when the tracker changes its sampling time as explained in the next section.

6.4.3 Obfuscation Scheme

The obfuscation privacy scheme combines the position perturbation scheme and the random beaconing rate together in order to avoid reducing the achieved distortion level using a different tracker sampling time. Formally, the beacon $\overline{B}_k$ sent at time k is modified as follows:

$$\overline{B}_k = \begin{cases} \textit{dropped} & k \bmod (t_r \cdot t_b) \neq 0 \\ \hat{B}_k & k \bmod (t_r \cdot t_b) = 0, t_{c(i)} < k \leq t_{c(i)} + d_n \cdot t_b \\ B_k & k \bmod (t_r \cdot t_b) = 0, t_{c(i)} + d_n \cdot t_b < k \leq t_{c(i+1)} \end{cases} \tag{6.2}$$

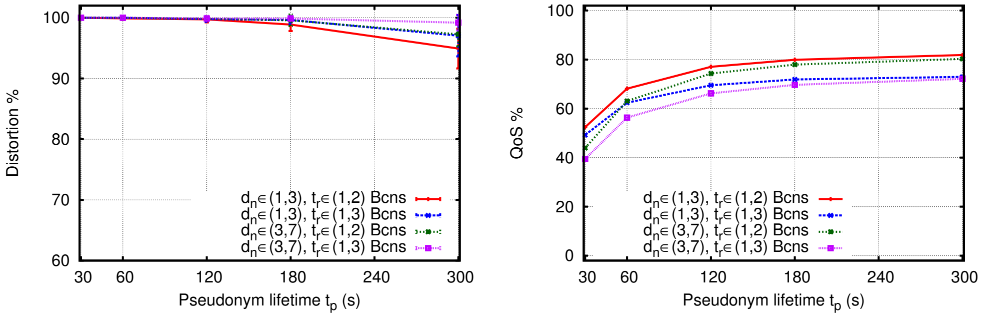
where $\hat{B}_k$ is the beacon of noised position $\hat{P}_k$ as defined in Equation 6.1.



(a) Privacy

(b) QoS

Figure 6.4: Privacy and QoS levels of the obfuscation scheme in a sparse traffic of 50 vehicles.



(a) Privacy

(b) QoS

Figure 6.5: Privacy and QoS levels of the obfuscation scheme in a dense traffic of 200 vehicles.

In Figure 6.4(a), the distortion of the obfuscation scheme with several combinations of random noise periods and random beaconing intervals is evaluated. In this experiment, a noise of 100 m is added to the vehicle positions for the specified random period after a pseudonym change. Two tracker sampling times of 0.5 and 7 s are employed and the average distortion is drawn in the figure. The obfuscation scheme results in complete distortion for the reconstructed tracks for $t_p \leq 180$ s regardless of the noise period, random beaconing interval and tracker sampling time. The distortion decreases slightly up to 90% for longer pseudonym lifetimes and according to the employed random noise periods and random beaconing intervals. The QoS is negatively affected especially when $t_p \leq 120$ s or $t_r \in (1, 3)$ beacons, as illustrated in Figure 6.4(b).

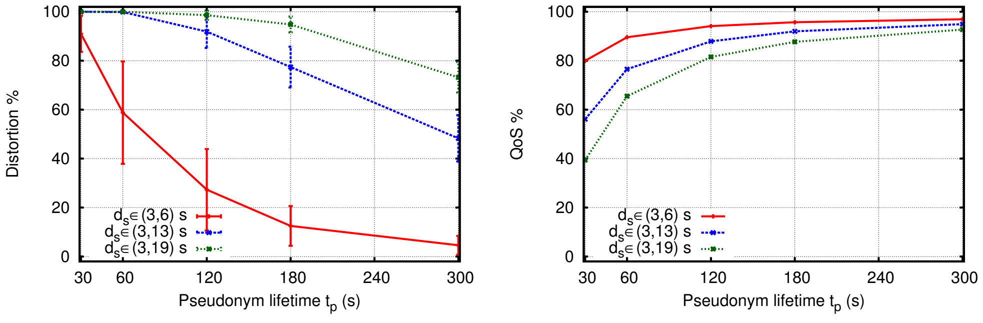
The QoS behavior of the obfuscation scheme follows the behavior of the position perturbation scheme but it is shifted down by the impact of the random beaconing interval (see Figures 6.1(b) and 6.3(b)). It is important to carefully choose the random beaconing interval because even a relatively short interval (e.g., $t_r \in (1, 3)$ beacons) results in reducing the number of broadcast beacons significantly per time step which in turn reduces the overall QoS. In Figure 6.5, the privacy and QoS of the obfuscation scheme are evaluated in the dense traffic of 200 vehicles of STRAW traces. As expected, the distortion is increased due to the increased vehicle density and thus the tracker is more confused in reconstructing correct tracks. The QoS is slightly increased due to the low speed of vehicles in the dense traffic.

According to these results, the obfuscation scheme offers significant confusion to the adversary which is reflected in complete distortion levels but at the cost of QoS. To achieve a QoS of 80%, beacons should be sent every random beaconing interval of (1, 2) beacons at most. Also, the pseudonym should be changed every 300 s followed by a random noise period of (1, 3) beacons. This configuration results in a distortion of 91%.

6.4.4 Comparison with Random Silent Period

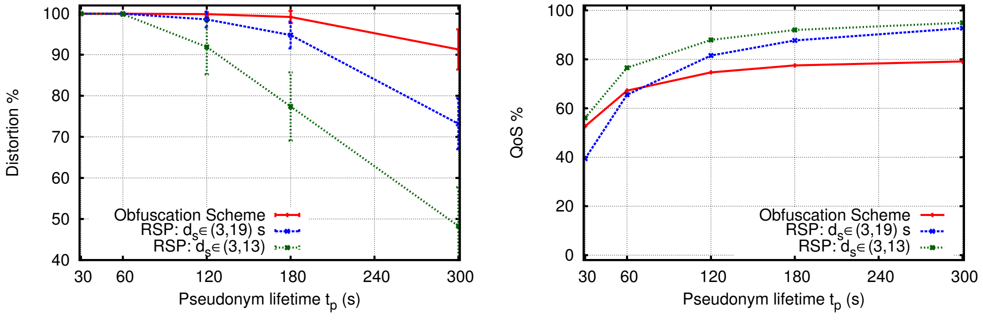
In this section, the obfuscation privacy scheme is compared with the random silent period (RSP) privacy scheme. The RSP lets a vehicle change its pseudonym after a fixed pseudonym time and keep silent for a uniformly random period within a range. The current American SAE J2735 standard [6] recommends keeping silent for 3 to 13 s or for duration of 50 to 250 m, which ever comes first after a pseudonym change. In Figure 6.6, the RSP is evaluated with several random periods and tracker sampling times between 0.5 s and the maximum allowed silence of each period. The RSP achieves a high distortion level when the silent period is sufficiently long ($d_s \in (3, 13)$ s or longer) and the pseudonym lifetime is relatively short ($t_p \leq 180$ s). The QoS is significantly reduced in short pseudonym lifetimes due to the frequent silence periods.

We selected a parameter set for the obfuscation scheme and compared it with the RSP. The selected obfuscation scheme parameters are position noise $\sigma = 100$ m, $d_n \in (1, 3)$ beacons and $t_r \in (1, 2)$ beacons which result in the highest QoS in the previous experiments. The RSP is evaluated with two silent periods (3, 13) and (3, 19), as shown in Figure 6.7. The obfuscation scheme achieves a higher distortion level than the RSP especially with long pseudonym lifetimes. However, the QoS of the obfuscation scheme is generally lower than the QoS of the RSP especially with long pseudonym lifetimes. This reduction in QoS results from the random beaconing interval which eliminates a reasonable number of beacons every time step. These eliminated beacons prohibit the in-vehicle



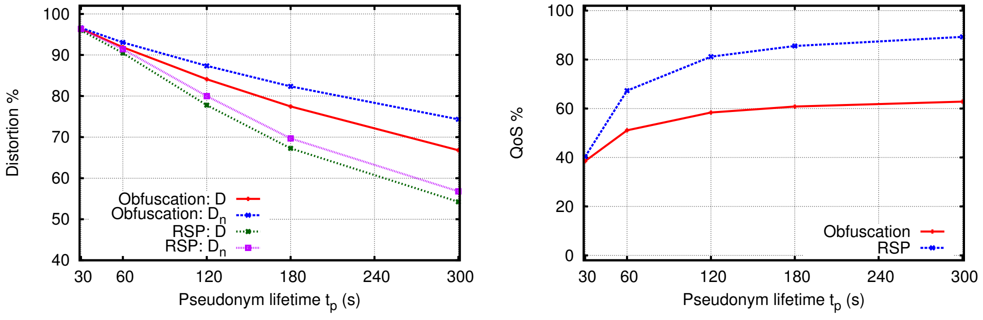
(a) Privacy (b) QoS

Figure 6.6: Privacy and QoS levels of the random silent period (RSP) scheme.



(a) Privacy (b) QoS

Figure 6.7: Comparison between the obfuscation privacy scheme and RSP in STRAW traces of 50 vehicles.



(a) Privacy (b) QoS

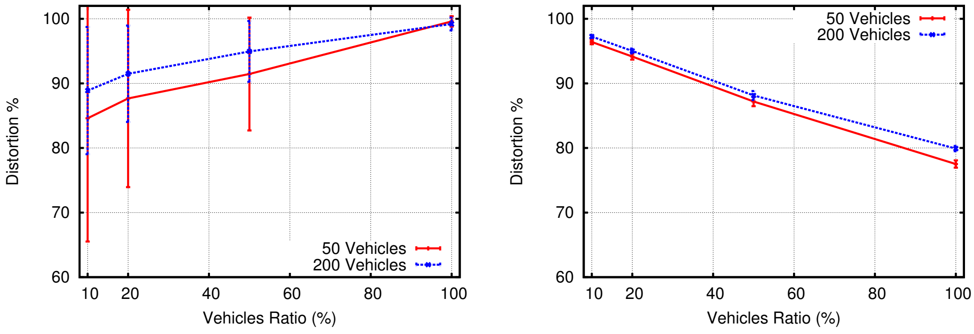
Figure 6.8: Comparison between the obfuscation privacy scheme and RSP in realistic traces.

tracker from estimating the states of other vehicles correctly. We repeat this experiment using realistic traces presented in Section 1.6.3. The same parameters are employed for the obfuscation scheme while the silent period of (3, 19) s is selected for the RSP. In Figure 6.8(a), the distortion (D) and normalized distortion (D_n) are shown for both schemes. In general, the distortion level decreases with the increase of the pseudonym lifetime t_p because the average lifetime of traces is not sufficiently long to allow several pseudonym changes when t_p is long. However, the obfuscation scheme results in a higher (normalized) distortion than RSP especially with long pseudonym lifetimes. The QoS of both schemes is shown in Figure 6.8(b). The QoS of the RSP does not differ from that in the STRAW traces and it is much higher than the QoS of the obfuscation scheme. The impact of the beacon elimination in the obfuscation scheme is increased with the 1 s time step of the realistic traces.

6.4.5 Partial Obfuscation

The results presented in previous sections confirm the effectiveness of the obfuscation scheme in preserving privacy but with a reasonably negative impact on the QoS, even worse than the RSP scheme. As a workaround, we evaluate the behavior of the obfuscation scheme when only a ratio of the vehicles are applying the obfuscation scheme while the remaining vehicles change their pseudonyms periodically with no other privacy mechanisms. This partial adoption of the obfuscation scheme shows the effectiveness of the scheme when only some drivers are concerning privacy while the majority of drivers concern the QoS of applications. In the following experiments, we apply the obfuscation scheme on a ratio of the vehicle traces and measure the distortion level of these traces. Nevertheless, the QoS is measured over the whole traces because the safety applications use information received from all vehicles regardless of applying the obfuscation scheme. The first experiment employs the STRAW traces of 50 and 200 vehicles. We repeat the experiment of each ratio 10 times where different traces are randomly selected for applying the obfuscation scheme each time. We employed the 10 variations of these vehicle densities which means that each vehicles ratio is evaluated 100 times.

Figure 6.9(a) illustrates the distortion level of various vehicle ratios applying the obfuscation scheme with $t_p = 180$ s, $\sigma = 100$ m, $d_n \in (1, 3)$ beacons and $t_r \in (1, 2)$ beacons. We applied tracking with several tracker sampling times t_s ranging from 0.5 s to 3 s. We notice that the average distortion level increases with the increase of the vehicle ratio with a large standard deviation represented in error bars. However, the average distortion is still high even for small vehicle ratios (e.g., an average distortion of 88% can be achieved for a vehicle ratio of 20%). In addition, the QoS is largely enhanced with small



(a) Privacy

(b) QoS

Figure 6.9: Privacy and QoS levels of partial obfuscation ratios in STRAW traces. ($t_p = 180$ s)

vehicle ratios (e.g., a QoS of 94% can be achieved for a vehicle ratio of 20%), as shown in Figure 6.9(b). The same experiment is repeated with the realistic traces, as shown in Figure 6.10. The vehicle ratio of 20% results in a normalized distortion of 73% and QoS of 87% which is a much better compromise between preserving privacy and providing high QoS when compared with the 100% vehicle ratio. According to these results, the obfuscation scheme can be employed to preserve location privacy without a reasonable impact on the QoS only if a small ratio (e.g., 20%) of vehicles are applying it.

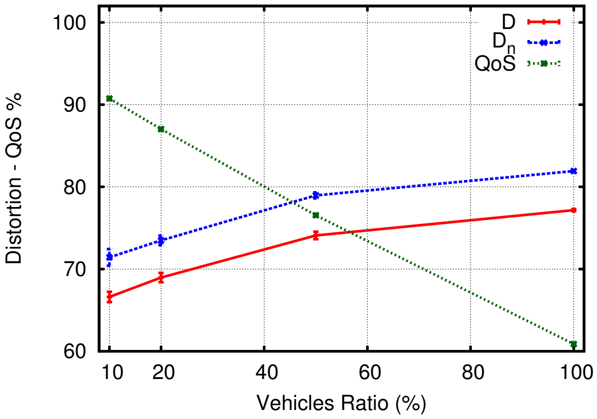


Figure 6.10: Privacy and QoS levels of partial obfuscation ratios in realistic traces. ($t_p = 180$ s)

6.5 Advanced Adversary

The employed adversary in previous experiments does not exploit the knowledge of the operations of the obfuscation scheme. An advanced adversary may exploit this knowledge and try to skip the noised beacons and complement the eliminated beacons. An example of advanced adversaries is illustrated in Figure 6.11.

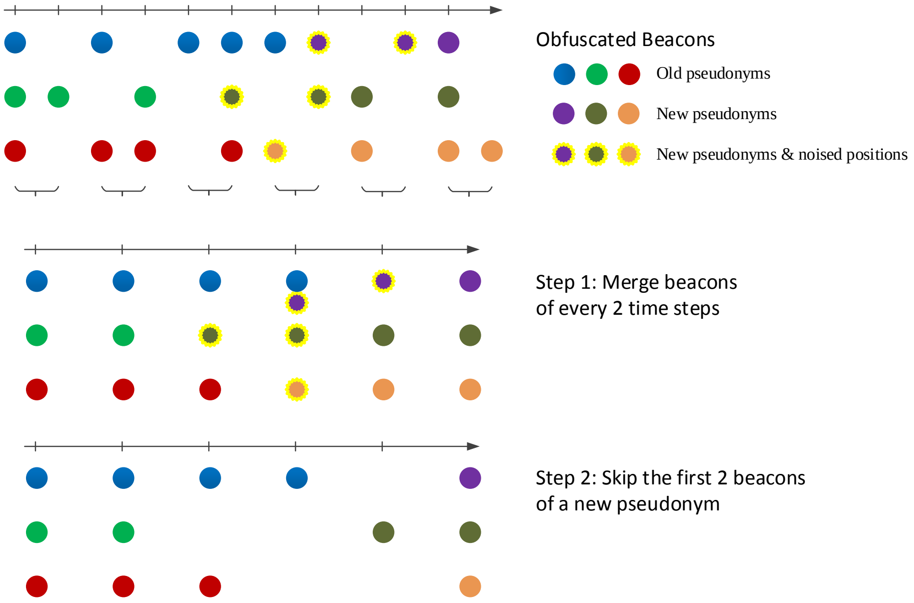
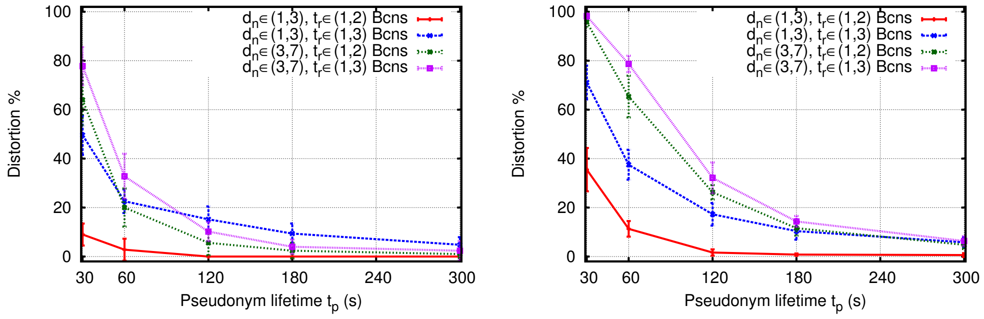

Obfuscated Beacons
Old pseudonyms
New pseudonyms
New pseudonyms & noised positions
Step 1: Merge beacons of every 2 time steps
Step 2: Skip the first 2 beacons of a new pseudonym

Figure 6.11: Advanced adversary for the obfuscation scheme.

In this example, there are three vehicles that apply the obfuscation scheme of $d_n \in (1, p)$ beacons and $t_r \in (1, q)$ beacons where $p = 3$ and $q = 2$ beacons. Their obfuscated beacons over 12 time steps are drawn as circles whose color refers to the vehicle pseudonym, as shown in the upper part of the figure. The advanced adversary performs two preprocessing steps on these beacons before tracking. Firstly, the adversary merges beacons of every q time steps into a single time step by averaging the beacon data (i.e., position and velocity) of the same pseudonym, as presented in the middle part of the figure. Secondly, the adversary eliminates the first $\lceil p/q \rceil$ beacons of a new pseudonym to skip the noised period after a pseudonym change, as illustrated in the bottom part of the Figure. This advanced adversary turns the obfuscation scheme into a silent period scheme with a silence period up to $\lceil p/q \rceil$ beacons after reducing the beaconing rate by $1/q$. Since p and q are expected to be small to minimize the impact on the QoS, tracking of the processed beacons can achieve a very



(a) 50 Vehicles

(b) 200 Vehicles

Figure 6.12: The obfuscation scheme evaluation against the advanced adversary.

high accuracy as shown next.

We evaluate the obfuscation scheme of $\sigma = 100$ m against this advanced adversary using the STRAW traces. Figure 6.12 shows the distortion level of various noise periods and random beaconing intervals in both sparse and dense traffic. The distortion level decreases significantly especially with the sparse traffic and pseudonym lifetimes $t_p > 30$ s. This result shows the ineffectiveness of both position perturbation and random beaconing against this adversary.

6.6 Summary

In this chapter, we investigated the applicability of obfuscation mechanisms in VANET. A combination of position perturbation and random beaconing interval is proposed and evaluated using both STRAW and realistic traces. The QoS is also evaluated showing the negative impact of the obfuscation scheme on safety applications. In addition, we come up with an advanced adversary who can overcome the operations of the obfuscation scheme and track vehicles with low distortion levels. The random beaconing rate is overcome by merging beacons of subsequent time steps while the position perturbation period is avoided by skipping beacons of new pseudonyms for the max noising period. According to these results, we can conclude that the obfuscation mechanisms not only have a significant negative impact on the QoS of safety applications but also are ineffective in preserving location privacy. Therefore, they should not be used in preserving the location privacy in VANET domain.

7 Context-based Privacy Schemes

7.1 Introduction

Since the obfuscation privacy scheme do not achieve reasonable privacy and QoS levels, other schemes are investigated. Elimination-based privacy schemes are commonly used in vehicular networks whether time-based such as the random silent period (RSP) or location-based such as mix-zone. The concept of these schemes is to hide vehicle information for a sufficient period of time before a pseudonym change. This discontinuity in the spatiotemporal information makes it more difficult to correlate beacons of the old and new pseudonyms. However, this approach may not be effective in certain cases. For example, in the random silent period scheme, if a vehicle switches to silence alone, then an adversary can track it because no other vehicle has changed its pseudonym. Even if several vehicles enter silence together, the adversary can still track them if their routes are distinguishable from each other and follow tracks predicted by the adversary. The RSP may also negatively affect the QoS of safety applications in relatively long silence, as discussed and illustrated in Section 5.4. Safety applications evaluate the surrounding traffic and provide information or warnings to the driver based on the data extracted from beacons received from other vehicles. Therefore, interruptions in these beacons due to silence periods may negatively affect the provided warnings.

In this chapter, we propose a context-aware privacy scheme (CAPS) that allows a vehicle to select the effective context in which to enter a silence period and change its pseudonym and when to resume beaconing with a high probability of confusion to a global adversary. This scheme monitors surrounding vehicles through their beacons using the vehicle tracker proposed in Section 3.4.7. The motivation behind using an in-vehicle tracker is to provide a more realistic view about the surrounding traffic and facilitate estimating the likelihood of confusion to an adversary. We evaluate this scheme using both the STRAW and realistic vehicle traces in comparison with the random silent period scheme. Moreover, we improve the CAPS by proposing a context-adaptive scheme (CADS) which minimizes the required parameters by adapting itself according to the vehicle context and the driver's privacy preference. Last but not least, several privacy schemes are evaluated and compared with our context-based schemes.

7.2 Related Work

Some context-based privacy schemes are proposed in VANET domain, as discussed in Section 2.6.2. Gerlach and Guttler [57] proposed the concept of a *context mix*, where a vehicle changes its pseudonym after holding the pseudonym for a stable time when there are N neighbors within a small radius . Lu et al. [87] proposed changing pseudonyms in social spots such as signaled intersections and parking areas where several vehicles remain stopped for some time. Before leaving a social spot, vehicles change their pseudonyms to create a mix zone. Some variations of silent period schemes that take safety into consideration were also proposed. Buttyán *et al.* [28] proposed ceasing to send messages when the vehicle moves slowly. The rationale for choosing a low speed is that silent periods are less likely to cause fatal accidents at low speeds and these low speeds indicate natural mixing areas, where many vehicles are in close proximity. Wei and Chen [139] proposed obfuscating the position, speed and direction within the safe distance radius calculated by a safety analysis algorithm. Additionally, they propose changing the length of the silent period based on the distance from other vehicles such that the closer the vehicles, the shorter the silent period.

Our context-based schemes differ from and improve the previously mentioned techniques. First, the proposed schemes do not rely exclusively on fixed heuristics, such as a changing velocity or a density threshold, to choose the appropriate situations to change pseudonyms. On the contrary, they monitor the vehicle context and decide dynamically when and where keep silent to change pseudonyms and when to resume beaconing. This dynamic context-based technique provides short but efficient silence periods so that the QoS of safety applications is not significantly affected. Second, our schemes conserve the pseudonyms pool of a vehicle by increasing the minimum pseudonym lifetime when pseudonyms are changed several times with likely tracker confusions. Third, the proposed schemes are evaluated using realistic large-scale vehicle traces which confirms their practicability, applicability and scalability in real-world situations.

7.3 System and Adversary Models

We assume the same system model, as defined in Section 1.5.1. For the adversary model, we consider protecting vehicles from both (1) a global passive adversary (GPA) and (2) a local active adversary (LAA). The GPA can monitor all exchanged messages, as defined in Section 1.5.2. The LAA can send authenticated messages to the network through a limited amount of compromised vehicles driving in the road network. It is assumed that it is extremely difficult for an active adversary to be global. The LAA aims mainly to deplete the pseudonym pools of the victim vehicles by forcing repeated pseudonym changes. If its pool is depleted, the victim will attempt to refill its pseudonym pool by initiating a pseudonyms issuing process with a trusted service provider, which is not always accessible. This adversary tries to mimic conditions that make its surrounding vehicles change their pseudonyms by exploiting the procedures of the privacy scheme. Since our proposed schemes depend on the vehicle context to change pseudonyms, it is important to evaluate them against active attacks. The encryption-based privacy schemes (such as CMIX [53]) fails in protecting vehicles from this adversary model because the compromised vehicles can obtain symmetric keys from RSUs and decrypt all exchanged messages. This gives another advantage for our proposed schemes.

7.4 Context-aware Privacy Scheme (CAPS)

7.4.1 CAPS Concept

The basic concept of our Context-aware Privacy Scheme (CAPS) is to determine the appropriate context in which a vehicle should change its pseudonym. This approach aims to increase the effectiveness of such changes against tracking and avoid wasting pseudonyms in easily traceable situations. More specifically, a vehicle continuously monitors other vehicles located within its communication range and tracks their beacons using an NNPDA tracker. As explained in Chapter 3, the NNPDA is an efficient multi-target tracking algorithm that has exhibited a high tracking accuracy for anonymous beacons with different amounts of noise and beaconing rates.

As illustrated in Figure 7.1, the CAPS works as follows. During its active status, the subject vehicle (SV) uses its current pseudonym in beacons until the pseudonym lifetime reaches a minimum time. Once it exceeds this time, the vehicle checks whether any of monitored *neighbors* missed its beacons for several time steps. Here, neighbors refer to vehicles located within a predefined radius from the subject vehicle (e.g., 50 or 100 m). If the SV finds a silent neighbor, it turns to silence as well. Otherwise, it continues using its current pseudonym until its lifetime reaches a maximum pseudonym time and then the vehicle turns to silence.

When a vehicle is silent, it returns to beaconing under more complex conditions based on the *gating* phase of vehicle tracking. It was explained in Section 3.2.2 that a gating process is required in target tracking to eliminate unlikely measurement-to-track associations from being tested. It requires any new mea-

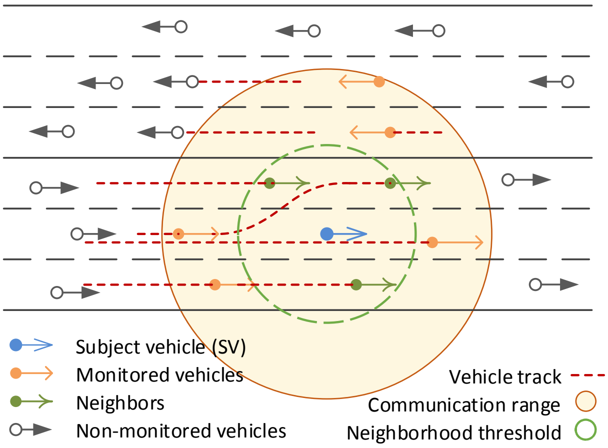


Figure 7.1: Illustration for the CAPS operations

surement to be located within the track *gate* to be a valid candidate for association with this track. The most common gating technique is ellipsoidal which defines the norm of the residual vector (d^2):

$$d^2 = \tilde{z}^T S^{-1} \tilde{z} \tag{7.1}$$

where $\tilde{z}$ and S are the residual vector and its covariance matrix obtained from the Kalman filter, respectively. We exploit this fact and require the beacon after silence to achieve one of the following two conditions to guarantee no correlation with previous beacons. As illustrated in Figure 7.2, the SV state should be nearer to the track of a silent neighbor than its original track or completely outside the gate of its original track. When these conditions hold, the adversary will most probably become confused when attempting to correlate this new beacon because it will not be assigned to its original track.

Formally, when the SV is silent, it continues monitoring surrounding vehicles and waits for the minimum silence time. Once exceeded, it checks if one of the following conditions holds regarding the norm of the residual vector (d^2) between its actual and estimated states:

1. $d^2 > d^2_{Nmin}$, where d^2_{Nmin} is the minimum norm of the residual vector between the SV actual state and the estimated states of its silent neighbors, as shown in the upper part of Figure 7.2.

2. $d^2 > max\_gate$, where max_gate is the maximum gate that the adversary may use, as shown in the lower part of Figure 7.2.

If one of these conditions holds, this new beacon is likely to be mixed with one of its silent neighbors or recognized as a new vehicle. Therefore, it is a

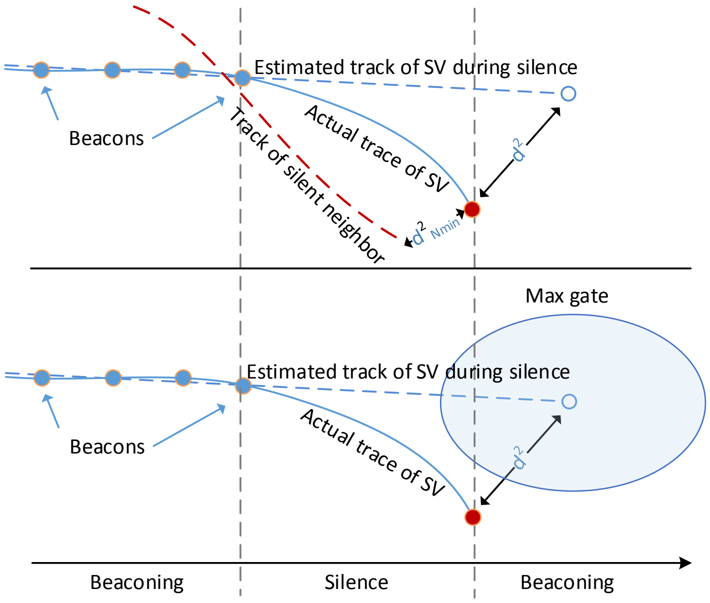


Figure 7.2: Illustration for the two conditions to exit silence.

suitable time to exit silence with a new pseudonym. If these conditions never occur, the SV remains silent until a maximum silence time is reached.

7.4.2 CAPS Algorithm

Algorithm 1 and its supporting functions presented in Algorithm 2 show the implementation details of the CAPS. Note that the SV uses this algorithm at every time step to decide on its next status whether active or silent. Algorithm 1 takes as input the tracks maintained for other vehicles ($other\_tracks$), a track for the SV itself ($myself\_track$), beacons received by the SV at the previous time step ($scan$), the current state of the SV obtained from its sensors ($actual\_state$) and its current status whether active or silent ($status$).

First, the tracks of the monitored vehicles are updated by the received beacons by calling the $update\_tracks$ function, which runs the NNPDA tracker to assign each beacon in a $scan$ to its corresponding track. Next, these tracks are stepped forward to the current time step by calling the $kalman\_predict$ function. Next, the candidates of silent neighbors are identified by calling the get_silent_cand function defined in Algorithm 2. This function finds neighbor tracks that are not updated by a beacon for at least the last $miss\_beacon\_threshold$ time steps and are located within the $neighborhood\_threshold$ from the SV. The $miss\_beacon\_threshold$ aims to discriminate between silent neighbors and

Algorithm 1 Context-Aware Privacy Scheme (CAPS)

```
Input: $other\_tracks, myself\_track, scan, actual\_state, status$
$update\_tracks(other\_tracks, scan)$;
$kalman\_predict(other\_tracks)$;
$sil\_cand := get\_silent\_cand(other\_tracks, actual\_state)$;
if $status = active$ then
    $psynm\_time := psynm\_time + 1$;
    if $psynm\_time > psynm\_max$ then
        $status := silent$;
    else if $psynm\_time > psynm\_min$ then
        if SIZE$(sil\_cand) >= sil\_node\_threshold$ and RAND$() > 0.5$ then
            $status := silent$;
        end if
    end if
    if $status = silent$ then
        $sil\_time := 1$;
    else
        $send\_beacon()$;
        $kalman\_update(myself\_track, actual\_state)$;
    end if
else                                                    ▷ $status = silent$
    $sil\_time := sil\_time + 1$;
    if $sil\_time \geq sil\_max$ then
        $state := active$;
    else if $sil\_time > sil\_min$ then
        if SIZE$(sil\_cand) > 0$ then
            $myself\_dist := calc\_dist(myself\_track, actual\_state)$;
            $min\_neigh\_dist := calc\_min\_dist(sil\_cand, actual\_state)$;
            if $(min\_neigh\_dist < myself\_dist$ or
                $myself\_dist > max\_possible\_gate)$ and RAND$() > 0.5$ then
                $status := active$;
            end if
        end if
    end if
    if $status = active$ then
        $psynm\_time := 0$;
        $psynm := get\_new\_pseudonym()$;
    end if
end if
$kalman\_predict(myself\_track)$;
```

neighbors whose beacons were missed due to a communication problem. The $neighborhood\_threshold$ aims to discriminate between silent neighbors and neighbors that tend to drive far from the SV. This threshold affects the tendency of switching to silence. If this threshold is large, then the SV is more likely to find silent neighbors, and thus, tends to switch to silence sooner and may consume more pseudonyms in dense traffic. If this threshold is small, then the SV tends to continue beaconing until a very close neighbor turns to silence.

In Line 8 of Algorithm 1, when the pseudonym lifetime ($psynm\_time$) exceeds the minimum pseudonym time ($psynm\_min$), the vehicle turns to silence when there are silent neighbor candidates more than $sil\_node\_threshold$. This threshold is generally set to one but it can be increased to protect against the LAA pseudonym depletion attack, as will be discussed in Section 7.5.4. We added randomization to the switching condition to prevent the adversary from guessing the exact time of turning to silence. In Line 16, if the SV does not turn to silence, it sends a beacon and updates its own track.

In Line 23, if the SV is already silent and its silence period exceeds the minimum silence time, it switches to active status if there are other silent neighbors and one of the following conditions holds. First, it calculates the norm of the residual vector between its actual state and its own track ($myself\_dist$) by calling the $calc\_dist$ function. It also calculates the minimum norm of residual vectors between its actual state and its silent neighbor tracks (min_neigh_dist) by calling the $calc\_min\_dist$ function. If the min_neigh_dist is less than the $myself\_dist$, then the actual state of the SV is most likely to be mixed with a track of those silent neighbors. Thus, the adversary would also become confused if this silent neighbor did not resume its beaconing in the same time step. Note that the vehicle track ($myself\_track$) simulates the adversary's knowledge about the SV because it is updated by the sent beacons only during active status and predicted during silence. Another condition for switching back to beaconing is that the $myself\_dist$ is larger than the maximum possible gate used by the adversary ($max\_possible\_gate$). In this case, the actual state of the SV is much farther than the state predicted by the adversary, and thus, a new track will be created for this new beacon. A randomization condition is also added to prevent the adversary from guessing the exact time of returning to beaconing. Once the SV exits silence, it uses a new pseudonym from its preloaded pool and resumes sending beacons at the next time step.

7.4.3 Experiment Results

In this section, we evaluate CAPS in terms of the distortion level and the QoS of safety applications. The tracker presented in Section 3.4.7 is employed to reconstruct traces from the beacons protected by CAPS. Unless specified otherwise,

Algorithm 2 Supporting Functions

```
function $get\_silent\_cand(other\_tracks, actual\_state)$
    $n :=$ SIZE$(other\_tracks)$;
    for $i := 1, n$ do
        if $other\_tracks(i).updated\_from \geq miss\_beacon\_threshold$ and
          EUCLID$(other\_tracks(i).pos, actual\_state.pos) <$
          $neighborhood\_threshold$ then
            $silent\_cand.ADD(other\_tracks(i))$;
        end if
    end for
    return $silent\_cand$;
end function
function $calc\_dist(track, state)$
    $\tilde{z} = state - track.H \cdot track.\hat{x}$;
    $d^2 = \tilde{z}^T track.S^{-1} \tilde{z}$;
    return $d^2$;
end function
function $calc\_min\_dist(tracks, state)$
    $n :=$ SIZE$(tracks)$;
    $minD := Inf$;
    for $i := 1, n$ do
        $d := calc\_dist(tracks(i), state)$;
        if $d < minD$ then
            $minD := d$;
        end if
    end for
    return $minD$;
end function
```

we employ the sparsest scenario (50 vehicles) of the STRAW vehicle traces described in Section 1.6.2 to generate pseudonymous beacons. This low density is selected because the tracker can track vehicles with low distortion levels even in the presence of intermediate random silent periods, as shown in Section 4.4. Thus, it will be a challenge for the CAPS to reduce tracking vulnerability in this low density scenario. Many experiments with different parameter combinations are performed using ranges specified in Table 7.1. The values of the minimum and maximum silence times and the maximum pseudonym time are guided by the European standard ETSI TS 102 867 recommendations [8]. In the next experiments, we show the effect of changing every other parameter while assigning the remaining parameters to their default values. Privacy is

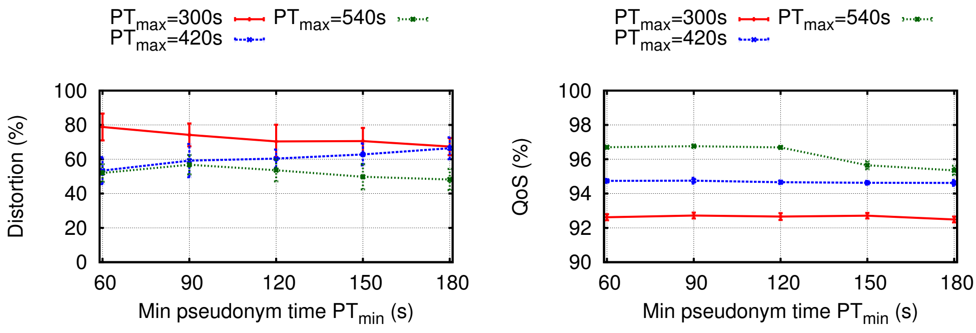
Table 7.1: CAPS parameter test ranges and default values

Parameter	Test Range	Default Value
Minimum pseudonym time PT$_{min}$ (s)	60 - 180	60
Maximum pseudonym time PT$_{max}$ (s)	300 - 540	300
Minimum silence time ST$_{min}$ (s)	0 - 5	3
Maximum silence time ST$_{max}$ (s)	7 - 13	13
Number of vehicles (V)	50 - 200	50
Neighborhood threshold (m)	10 - 100	50
Packet delivery ratio PDR	0.6 - 1	1
Missed beacons threshold (Bcn)	1 - 7	3

measured in terms of the distortion metrics defined in Equations 4.10 and 4.11 while the QoS of safety applications is measured as specified in Equation 5.17.

First, we study the effect of the minimum and maximum pseudonym times. In Figure 7.3(a), the variation of the minimum and maximum pseudonym times versus the distortion metric is presented. We notice that the minimum pseudonym time has little effect compared to the maximum time. For a longer maximum pseudonym time, the CAPS has a longer time allowance to find other silent vehicles and thus tends to keep the same pseudonym. This behavior reduces the number of pseudonym changes and the accompanying silence periods in the sparse environment we use. The decrease of the pseudonym changes reduces the adversary confusions and the distortion level. The impact of the pseudonym period on the QoS is displayed in Figure 7.3(b). The QoS increases with longer maximum pseudonym times due to the decrease of the number of pseudonym changes which helps the in-vehicle tracker to better estimate the safety application requirements.

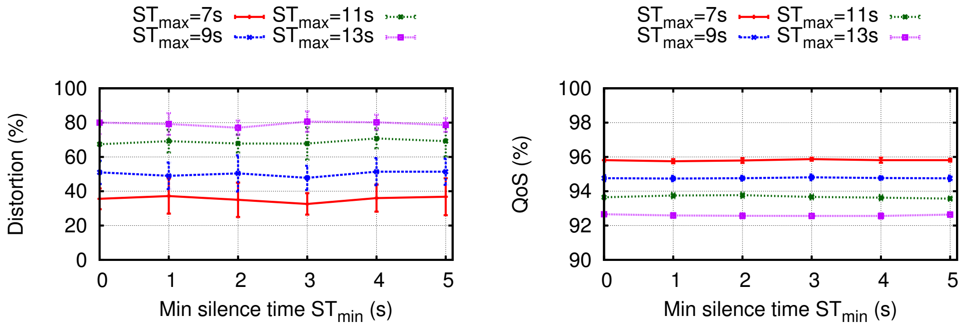
In Figure 7.4(a), we show the variation of the minimum and maximum silence times versus the distortion level. We notice again that the minimum silence time has little effect compared to the maximum time. Long maximum silence times (11 s and longer) give the context monitoring module the opportunity to find another silent neighbor with a track closer to the actual state of the subject vehicle than its own track. This opportunity results in an effective pseudonym change and highly probable confusion for the adversary reflected in the high distortion levels for maximum silence times of 11 s and longer. The QoS slightly decreases with the increase of the maximum silence time with almost no effect of the minimum silence time, as illustrated in Figure 7.4(b).



(a) Privacy

(b) QoS

Figure 7.3: Privacy and QoS levels of CAPS in several pseudonym periods.[1]



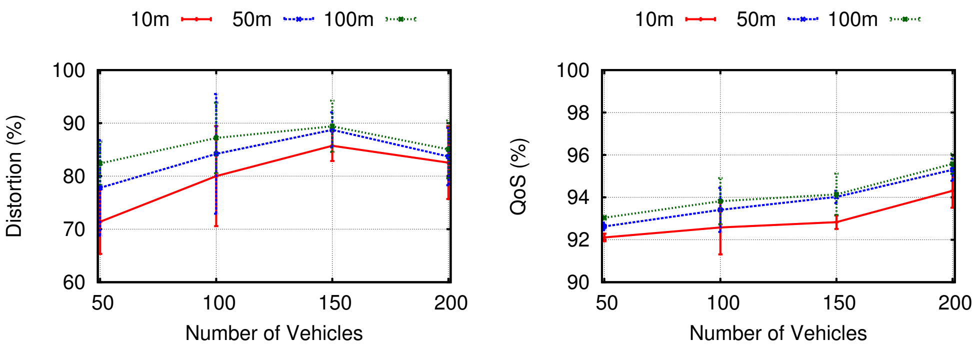
(a) Privacy

(b) QoS

Figure 7.4: Privacy and QoS levels of CAPS in several silence periods.

Furthermore, the effect of the neighborhood threshold and the vehicle density is evaluated in Figure 7.5. It can be noticed that large neighborhood thresholds are more effective than narrow ones, especially at lower densities, because a large threshold (i.e., large neighborhood circle) allows a vehicle to change its pseudonym sooner, as it is more likely to find another silent neighbor within this large circle. In dense environments, vehicles are already close to each other, and both narrow and large thresholds provide a sufficient area to find a silent neighbor, with the larger threshold having a slight advantage. The QoS is generally increasing with denser traffic because the speed error is lower in denser traffic. Note that the speed error is assumed to be 2% of the vehicle speed, as

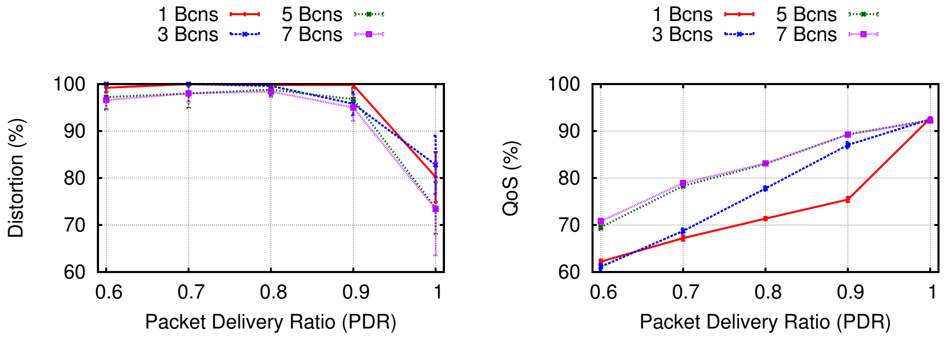
[1] The error bars in all figures represent the standard deviation.



(a) Privacy

(b) QoS

Figure 7.5: Privacy and QoS levels of CAPS in several neighborhood thresholds and vehicle densities.



(a) Privacy

(b) QoS

Figure 7.6: Privacy and QoS levels of CAPS in various missed beacon thresholds and packet delivery ratios.

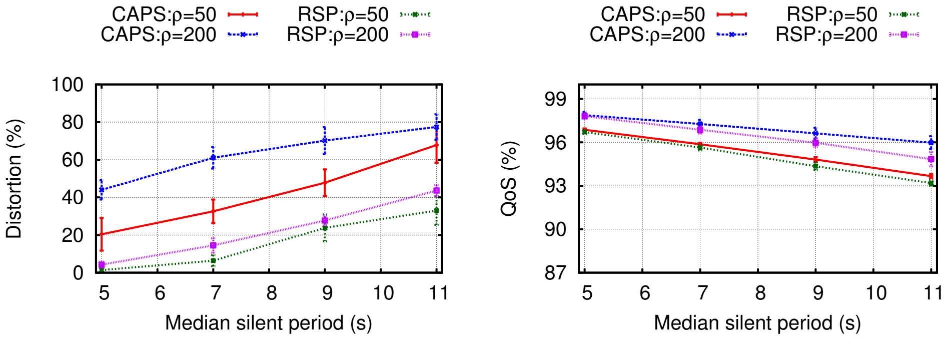
explained in Section 5.3.1 and the average speed in the dense traffic is lower than the sparse traffic.

The effect of the threshold of missed consecutive beacons for identifying silent neighbors in several packet delivery ratios (PDR) is investigated in Figure 7.6. It can be observed that when the PDR is less than one, the distortion increase significantly regardless the missed beacons threshold because of the large amount of beacon messages that the adversary failed to collect. In the perfect delivery case (i.e., PDR = 1), the missed beacons threshold affects the achieved distortion level. Small threshold values (i.e., ≤ 3) achieved a higher distortion level than large values. A small missed beacons threshold leads a vehicle to turn to silence based on weak evidence of the actual status of the neighbor. These last two observations regarding neighborhood and missed beacon thresholds are very important for context modeling. We initially thought that monitoring closer confirmed-silent neighbors would be the most essential, as these neighbors are the candidates for confusing the adversary. However, according to these findings, this confirmed-silent monitoring is less important than switching to silence sooner and letting the scheme wait for likely confusion during the silence period. As demonstrated in Figure 7.6(b), the QoS is the same for the all tested thresholds when the PDR equals one. Although the QoS generally decreases with lower PDRs, the missed beacons thresholds less than or equal the minimum silence time reduce the QoS significantly. This reduction results from the increased number of pseudonym changes the CAPS performs when the threshold is small. This increased number of pseudonym changes prevents the in-vehicle tracker from estimating the safety application requirements correctly.

Table 7.2: Parameters and results of the CAPS and RSP in STRAW traces for density of 50 vehicles

Parameters	Max silent time (s)	CAPS	5	7	9	11
		RSP	7	9	11	15
	Min silent time (s)	Both	3			
	Max pseudonym time (s)	CAPS	300			
	Fixed pseudonym time (s)	RSP	300			
	Min pseudonym time (s)	CAPS	60			
Results	Med pseudonym time (s)	CAPS	297	297	297	296
		RSP	294	291	288	286
	Med silent time (s)	CAPS	5	7	9	11
		RSP	5	7	9	10.8
	Pseudonym changes	CAPS	150	150	150	150
		RSP	165	165	165	165
	Confusion/Psynm Change	CAPS	0.14	0.22	0.32	0.48
		RSP	0.01	0.03	0.13	0.23

In addition, we compare the CAPS with the random silent period (RSP) scheme [115]. The RSP allows a vehicle to change its pseudonym after a fixed pseudonym time and keep silent for a uniformly random period within a preset range (e.g., from 3 to 13 s). As the two schemes have different assumptions and parameters, they are aligned based on the median silent and pseudonym times for all vehicles, actually performed in the simulation. In other words, we tried several values for the parameters of the RSP and obtained the resulting median silence and pseudonym times. We then compared these values with those obtained from the CAPS and matched the corresponding parameters. In Table 7.2, we show parameters passed to the CAPS and RSP that result in similar median silence and pseudonym times in STRAW traces of 50 vehicles. For example, using 5 and 7 s as maximum silence times in the CAPS and RSP, respectively, and 300 s as the maximum pseudonym time in both results in median silence times of 5 s and median pseudonym times of 297 and 294 s in the CAPS and RSP, respectively. Thus, when comparing these schemes, we use the corresponding parameter pairs shown in the first two rows in Table 7.2.

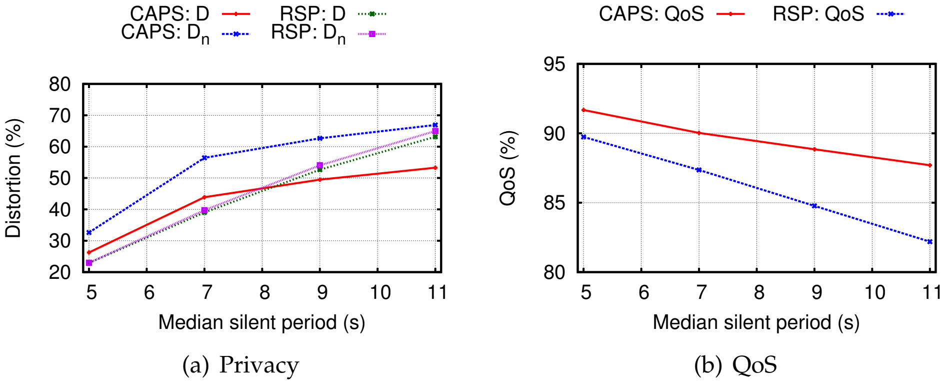


(a) Privacy

(b) QoS

Figure 7.7: Privacy and QoS levels of CAPS compared to RSP in STRAW traces of sparse and dense traffic.

As shown in Figure 7.7(a), the CAPS significantly increases the distortion level than the RSP does, especially for dense environments. In addition, the CAPS can achieve this higher distortion level using fewer pseudonyms. We added the number of pseudonyms used by all vehicles in the 50 vehicles scenario in the third row of the results section of Table 7.2. It is clear that the CAPS uses fewer pseudonyms. Furthermore, we measured the ratio of the adversary confusions per pseudonym change for both schemes to infer the effectiveness of the pseudonym change. According to ratios presented in the last row in the results section of Table 7.2, the pseudonym changes, that the CAPS performs, result in adversary confusions at least twice the confusions resulting from the pseudonym changes of the RSP. These last two results are important because they emphasize that choosing the situations in which to change pseudonyms



(a) Privacy

(b) QoS

Figure 7.8: Privacy and QoS levels of CAPS compared with RSP in realistic traces.

and keep silence is an effective and efficient way to preserve location privacy. Although the CAPS achieves higher distortion levels, it results in a higher QoS in safety applications than the RSP, as demonstrated in Figure 7.7(b). The QoS reduction of the RSP results from the increased number of pseudonym changes which are followed by silence periods. In relatively long silence periods, the in-vehicle tracker cannot estimate the states of the nearby vehicles correctly [2].

We repeat the previous experiment using realistic traces to confirm the applicability and effectiveness of the CAPS in real-world situations. As the distribution of the realistic traces is dynamic and different from that of the simulated traces, we found that the parameters, shown in Table 7.2, produce different median pseudonym times but similar median silence times. Thus, we used a fixed pseudonym time of 120 s instead of 300 s in the RSP to achieve alignment in both times, as shown in Table 7.3. Additionally, due to the huge number of traces, we ran this experiment once for each median silence time. As shown in Figure 7.8(a), the CAPS increases the distortion level (D) more than that of the RSP on average for relatively short median silence times ($\leq$ 7 s). However, the CAPS significantly increases the normalized distortion (D_n) than that of the RSP. The normalized distortion metric only considers the vehicles that have changed their pseudonyms during the simulation. Due to the long maximum pseudonym time of the CAPS, many vehicles have never changed their pseudonyms which results in zero distortion for their traces. The operations of both schemes have been further analyzed in different aspects, as presented in Table 7.3. We observe that the number of pseudonym changes made by the

[2] The distortion and QoS levels of the RSP seem different from that presented in Section 6.4.4 because the employed tracker in that section is not tuned to accommodate silence periods.

RSP are on average 1.6 times the number made by the CAPS, as shown in the third row of the results section of Table 7.3. This result indicates the efficiency of the CAPS in increasing the normalized distortion with fewer pseudonyms. Furthermore, we noticed that the increase in the normalized distortion level achieved by the CAPS is caused by the effectiveness of the pseudonym changes rather than their number. We observed that the ratio of adversary confusion per pseudonym change using the CAPS is 1.5-2.4 times greater than using the RSP depending on the length of the silence period, as shown in the last row of Table 7.3. This finding confirms the CAPS ability to choose the appropriate context for changing pseudonyms.

Table 7.3: Parameters and results of the CAPS and RSP in realistic traces

Parameters	Max silent time (s)	CAPS	5	7	9	11
		RSP	7	9	11	15
	Min silent time (s)	Both	3			
	Max pseudonym time (s)	CAPS	300			
	Fixed pseudonym time (s)	RSP	120			
	Min pseudonym time (s)	CAPS	60			
Results	Med pseudonym time (s)	CAPS	114	112	112	111
		RSP	114	111	108	105
	Med silent time (s)	Both	5	7	9	11
	Pseudonym change/vehicle	CAPS	1.50	1.49	1.48	1.48
		RSP	2.44	2.42	2.41	2.39
	Confusion/Psynm change	CAPS	0.26	0.55	0.64	0.71
		RSP	0.12	0.23	0.36	0.46

Regarding the QoS, we observe that both schemes have a lower QoS than that shown in the STRAW traces, but the CAPS still achieves an acceptable QoS of at least 88%, as shown in Figure 7.8(b). There are two issues that may explain this result. First, the time step of the realistic traces is 1 s but 0.5 s in the STRAW traces. This longer time step prevents the in-vehicle tracker from obtaining the desired accuracy especially in estimating the speed. We verified this finding by testing the STRAW traces again but skipping every other sample to produce a 1 s time step. The QoS of 1 s time step in the 11 s median silent period is 90% for both schemes, while it was 93% in the case of a 0.5 s time step. Second, the RSP changed pseudonyms much more often than the CAPS did, as shown in Table 7.3. As pseudonym changes are preceded by silence, the in-vehicle tracker failed to estimate the state of silent vehicles in the RSP more than in the CAPS, which is reflected in the lower QoS for RSP, especially in longer silence periods. This behavior is less noticeable in the STRAW traces because the increase in pseudonym changes of the RSP over the CAPS was not significant in the STRAW traces.

Based on these results, we can summarize the following findings. The CAPS achieves up to 35% increase in the distortion level on average from that of the RSP in a sparse environment of 50 vehicles where similar median silence and pseudonym times are used. This increase may reach up to 45% in a denser environment of 200 vehicles. However, the results show that the distortion level of CAPS is not sufficiently high and may allow vehicle tracking, especially with realistic traces. This may occur because of the relatively short length of traces (15 min in STRAW traces while 5 min in realistic traces). Longer traces may allow several pseudonym changes and tracker confusions. On the other hand, the CAPS achieves a better QoS of safety applications than the RSP does in realistic traces. In general, the impact of the CAPS on safety applications is not particularly significant, especially when short beaconing times are used (e.g., 0.5 s).

7.4.4 CAPS Efficiency

Regarding the efficiency of the CAPS, we implemented it using MATLAB as a centralized program, which operates on samples located in the communication range of each vehicle separately. We exploit the parallel for loop feature in MATLAB to iterate on vehicles asynchronously at every time step. We run our experiments on an Intel QuadCore i7-4800MQ @ 2.70GHz Hyper-threaded CPU. We calculate the running time of the CAPS to process samples received by a vehicle in a single time step and average over all vehicles and time steps. We found that the average running time is 5 ms for realistic traces. Note that this running time is obtained using a single thread, as the CAPS code is basically sequential. Thus, this running time is reproducible on single-thread single-core CPUs of the given speed. Therefore, we can conclude that the CAPS is efficient when high-end CPUs are used because the most frequent beaconing rate is 100 ms and the vehicle will have plenty of time to do other tasks. However, if lower-end CPUs are used in vehicles, then further code optimization should be investigated. The memory is not an issue, as the CAPS uses only a few hundreds of kBs for the Kalman filter tracks of the nearby vehicles.

7.4.5 CAPS Shortcomings

We note three shortcomings of the CAPS. First, we observe that some vehicles change pseudonyms unnecessarily several times with no significant advantage in increasing the distortion. Having a few confusions per trace is sufficient to avoid continuous vehicle tracking. However, frequent pseudonym changes and confusions may negatively affect the QoS of a safety application, as neighbors cannot estimate the vehicle state correctly. Therefore, we propose increasing the minimum pseudonym time each time a vehicle changes its pseudonym with a probable confusion. Second, the CAPS takes several parameters that may not be optimized for different traffic densities and situations. For example, a wide neighborhood threshold may be more suitable for sparse traffic than dense traffic. Third, the CAPS does not consider the driver's preference regarding privacy. In fact, privacy depends on the preferences of the user and the technical solutions should be adaptable to empower users to determine what is allowed with their personal information [17]. For example, it may be desirable to maximize the privacy level when the driver goes to a sensitive place. For these reasons, we propose a more advanced scheme that considers these shortcomings, which we call the context-adaptive privacy scheme (CADS) as explained next.

7.5 Context-adaptive Privacy Scheme (CADS)

The CADS allows a driver to choose among privacy preferences, whether low, normal or high. It optimizes the internal parameters dynamically according to the density of the surrounding traffic and the driver's privacy preference. In addition, it preserves the vehicle pseudonyms pool for a longer time if the pseudonym is already changed with a probable confusion.

To optimize the scheme parameters with respect to the surrounding traffic, we investigate the performance of the CAPS in two different densities; sparse and dense traffic. First, we select two relatively short sub-datasets from the realistic vehicle traces with low and high traffic densities. We then test the CAPS on each sub-dataset with many parameter combinations and obtain the resulting distortion and QoS metrics. Second, to incorporate the privacy preference in CADS, we divide the results of the sub-dataset experiments into three categories according to the achievable distortion. Next, we identify the parameters that result in the best compromise between distortion and QoS in each category. Third, we insert these categorized parameters of each density into CADS and bind them according to the real-time vehicle density and the input privacy preference. We next discuss each step with its accompanying experiments in

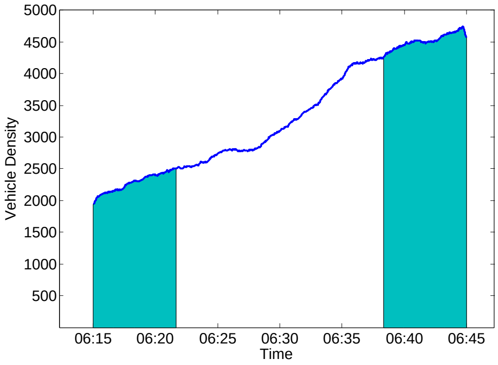


Figure 7.9: Vehicle density of realistic traces with sub-datasets highlighted.

detail.

7.5.1 Sub-datasets Evaluation

As explained in Section 1.6.3, the realistic traces have an increasing density range from 1,929 to 4,572 vehicles. We selected two sub-datasets, each 6 min long from the beginning and the end of the vehicle traces, as shaded in Figure 7.9. We excluded traces that last less than one minute from these sub-datasets. The CAPS is then evaluated using each sub-dataset and the following parameter combinations: maximum pseudonym times of 180, 240 and 300 s, maximum silence times of 7, 9, 11 and 13 s, neighborhood threshold of 50 and 100 m and increments of the minimum pseudonym time after a probable confusion of 0 or 60 s. We run the CAPS using these parameter combinations on both sub-datasets and obtain the achieved privacy and QoS metrics.

7.5.2 Parameters Selection

From all experiments tested in the previous step, we exclude those results with a QoS less than 85% as we assume that the safety application will not operate with an acceptable accuracy in such cases. Although the distortion and the QoS are inversely proportional, we notice that the QoS varies much less than the distortion. Therefore, the results are categorized based on the QoS instead, to facilitate categorization. The results are divided into low, normal and high privacy levels when they achieve the maximum, average and minimum QoS, respectively in each sub-dataset. Thus, the parameters for a high privacy preference are selected when a QoS of 85% is attained. The parameters for a low privacy preference are selected when the highest QoS is obtained but with a distortion of at least 25%. This low distortion constraint is added to ensure

Table 7.4: Optimized CADS parameters and their results

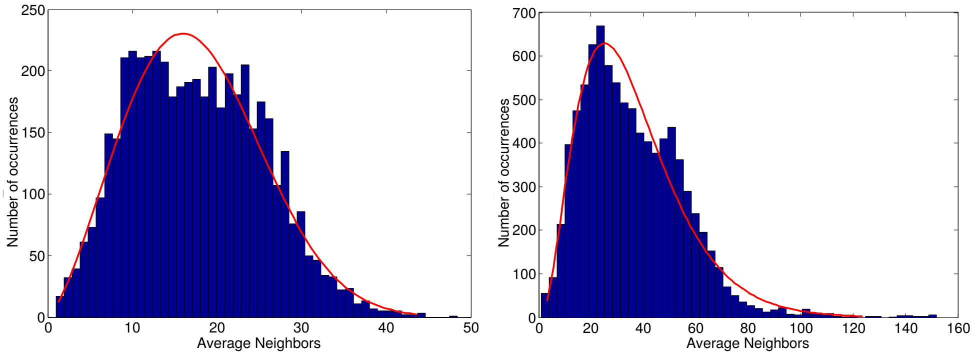
		Privacy Preference		
Parameter/Result	Density	Low	Normal	High
Max pseudonym time (s)	Sparse	240	300	180
	Dense	240	180	180
Max silence time (s)	Sparse	11	11	11
	Dense	11	13	11
Pseudonym time increment (s)	Sparse	60	60	0
	Dense	60	60	0
Neighborhood threshold (m)	Sparse	50	100	100
	Dense	50	50	100
Distortion (%)	Sparse	26	37	48
	Dense	31	46	55
Normalized Distortion (%)	Sparse	35	49	56
	Dense	44	57	65
QoS (%)	Sparse	90	87	85
	Dense	91	88	85

some privacy even when low privacy preference is selected. The parameters for normal privacy preference are selected when the average QoS is attained with the highest distortion.

In Table 7.4, we show the selected parameter set for each privacy preference and vehicle density. In the last three rows, we include the resulting distortion and QoS of each parameter set when applied to the sub-datasets. We notice that the achievable distortion in the sparse sub-dataset is lower than that achievable in the dense sub-dataset. The distortion can be increased using more restrict parameters but only at the cost of the QoS.

7.5.3 CADS Algorithm

The parameter table 7.4 is integrated into the CADS to let a vehicle choose the adequate parameter set based on the driver's privacy preference and the real-time density of the surrounding traffic. A vehicle can estimate the traffic den-



(a) Sparse sub-dataset

(b) Dense sub-dataset

Figure 7.10: Average number of neighbors encountered by a vehicle in both sub-datasets.

sity by evaluating the average number of neighbors encountered over time. For this purpose, we analyzed the distribution of neighbors in both sub-datasets, as shown in Figure 7.10. We notice that the average number of neighbors that a vehicle encounters is 30 and 68 with 95% confidence in the sparse and dense sub-datasets, respectively. Therefore, a neighbors threshold of 30 vehicles is assigned to discriminate between sparse and dense traffic. In other words, a vehicle continuously counts the surrounding vehicles in its communication range and calculates the average over time. If the average number of surrounding vehicles is lower than 30 then the traffic is considered sparse, otherwise it is considered dense.

The CADS pseudocode is presented in Algorithm 3. It is similar to the CAPS code along with some modifications. It additionally takes the driver's privacy preference ($priv\_pref$) and the parameter lookup table (PLT). In Line 3, the vehicle updates the average number of neighbors (avg_neig) encountered over time steps. Upon status switching from silent to active or vice versa, the vehicle looks up the parameter table PLT using the avg_neig and $priv\_pref$ to obtain the optimized parameter set for their values. In Line 17, the vehicle updates the maximum silence time (cur_sil_max) by the preset value in PLT. Similarly, in Line 35, if the vehicle switches to the active status upon a likely confusion, the vehicle increases the minimum pseudonym time ($psynm\_min$) by the pseudonym time increment obtained from PLT. It is worthy to note that the minimum pseudonym time is only increased if the silence period led to a probable confusion.

Algorithm 3 Context-Adaptive Privacy Scheme (CADS)

```
Input: $other\_tracks, myself\_track, scan, actual\_state, status, priv\_pref, PLT$
$update\_tracks(other\_tracks, scan)$;
$kalman\_predict(other\_tracks)$;
$avg\_neig := (avg\_neig * neigt + SIZE(other\_tracks))/(neigt + 1)$;
$neigt := neigt + 1$;
$sil\_cand := get\_silent\_cand(other\_tracks, actual\_state)$;
if $status = active$ then
    $psynm\_time := psynm\_time + 1$;
    if $psynm\_time > cur\_psynm\_max$ then
        $status := silent$;
    else if $psynm\_time > cur\_psynm\_min$ then
        if SIZE$(sil\_cand) >= sil\_node\_threshold$ and RAND$() > 0.5$ then
            $status := silent$;
        end if
    end if
    if $status = silent$ then
        $sil\_time := 1$;
        $cur\_sil\_max := PLT[''max\_sil'', avg\_neig, priv\_pref]$;
        $neigt := 1$;
    else
        $send\_beacon()$;
        $kalman\_update(myself\_track, actual\_state)$;
    end if
else                                                    ▷ $status = silent$
    $sil\_time := sil\_time + 1$;
    if $sil\_time \geq cur\_sil\_max$ then
        $state := active$;
    else if $sil\_time > sil\_min$ then
        if SIZE$(sil\_cand) > 0$ then
            $myself\_dist := calc\_dist(myself\_track, actual\_state)$;
            $min\_neigh\_dist := calc\_min\_dist(sil\_cand, actual\_state)$;
            if $(min\_neigh\_dist < myself\_dist$ or
                $myself\_dist > max\_possible\_gate)$ and RAND$() > 0.5$ then
                $status := active$;
                $cur\_psynm\_min := cur\_psynm\_min+$
                            $PLT[''psynm\_inc'', avg\_neig, priv\_pref]$;
            end if
        end if
    end if
    **if** $status = active$ **then**
        $psynm\_time := 0$;
        $psynm := get\_new\_pseudonym()$;
        $cur\_psynm\_max := PLT[''max\_psynm'', avg\_neig, priv\_pref]$;
        $cur\_neighborhood\_thershold := PLT[''neigh\_thre'', avg\_neig, priv\_pref]$;
        **if** $cur\_psynm\_min >= cur\_psynm\_max$ **then**
            $cur\_psynm\_min := cur\_psynm\_min - 30$;
        **end if**
    **end if**
**end if**
$kalman\_predict(myself\_track)$;
```

7.5.4 Experiment Results

The CADS is evaluated against two adversary models: GPA and LAA, as defined in Section 7.3. We consider the CADS distortion and QoS levels in the GPA experiments while we concern the pseudonym lifetime in the LAA experiments.

Location Privacy under GPA

We evaluated the CADS using realistic traces in two different scenarios. In the first scenario, all drivers select the same privacy preference, whether low, normal or high. In Figure 7.11, we show the distortion, the normalized distortion and the QoS of each privacy level as a bar chart. As a comparison with CAPS, these metrics are displayed as dashed lines when a maximum silent time of 11 s is set in CAPS.

The distortion and normalized distortion of CADS increases when drivers select a higher privacy preference with a slight decrease in the QoS. Compared to CAPS, the CADS achieves a better compromise between distortion and QoS. Specifically, when a high privacy preference is used, the CADS achieves a 15% higher distortion, a 8% higher normalized distortion but with a slight decrease in QoS (only 4%). When a low privacy preference is used, the QoS is enhanced by 2% while the normalized distortion is still more than 50%. In normal privacy preference, distortion is increased because of the adaptation of the parameters based on the traffic density. These results confirm the validity and effectiveness of the context-adaptability to find a practical compromise between privacy preference and QoS.

In the second scenario, we allow vehicles to select the preferred privacy level randomly based on given percentages. In this scenario, we aim to confirm that

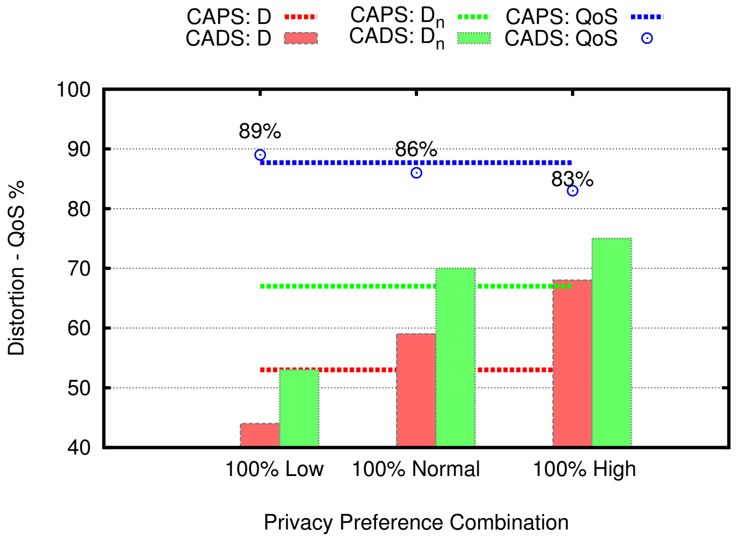


Figure 7.11: Comparison of the CADS evaluation when all vehicles use the same privacy preference and the CAPS evaluation with 11 s max silent time.

the privacy is more enhanced for vehicles that select a higher privacy level than the others. As the vehicles use a mix of privacy preferences, each privacy preference group is evaluated separately showing its distortion and normalized distortion. However, the QoS is evaluated over all vehicles, as lower-quality information obtained from vehicles that use a high privacy preference will affect other vehicles of lower privacy preferences and vice versa. In this scenario, we repeat each experiment five times with random selection of the privacy preference assigned to vehicles.

In the first and second experiments, 25% and 75% of vehicles use the normal privacy preference, respectively, while the rest uses the high privacy preference, as shown in Figure 7.12. Although both experiments employ swapped percentages of normal and high privacy levels, they achieve similar (normalized) distortion for both level groups with slight effect of the major group on the performance of the minor group.

In the third and fourth experiments, 75% of vehicles use the low privacy preference while the rest use normal and high levels, respectively. It is observable that the high level group in the fourth experiment achieves a higher distortion than that is achieved by the normal level group in the third experiment. Additionally, we notice that the high level group in the fourth experiment achieves slightly lower distortion than the same group in the second experiment. This

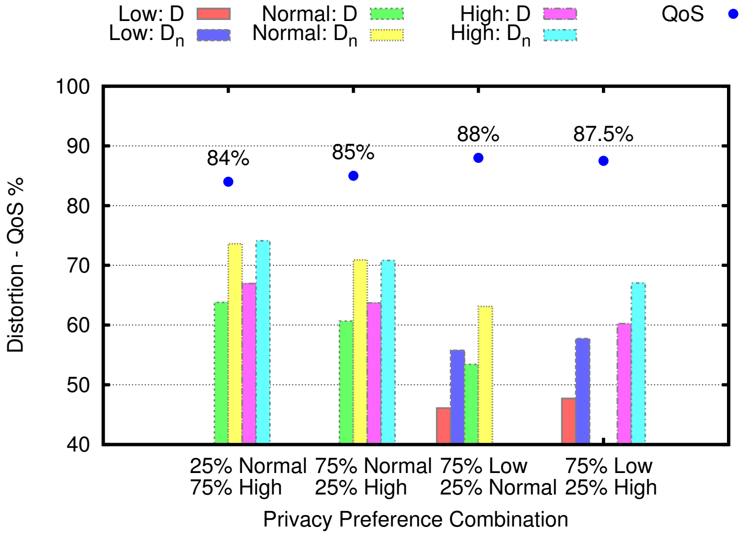


Figure 7.12: CADS evaluation when vehicles use a random privacy preference based on the specified percentages.

result may attributed to the major privacy preference group being low-level in the fourth experiment but normal-level in the second. Regarding the QoS, we notice that it follows the QoS of the major group with a slight effect from the minor. For example, the QoS in the first experiment is higher 1% than that in the "100% high-privacy" experiment, and the QoS in the fourth experiment is lower 1.5% than that in the "100% low-privacy" experiment. From all these observations, we can conclude that the distortion is mainly affected by the configured privacy level with a slight effect from the surrounding traffic. However, this change in distortion is compensated in the QoS.

Location Privacy under LAA

The local active adversary (LAA) performs a pseudonyms depletion attack which tries to force victim vehicles to change pseudonyms as soon as possible. It is important to evaluate context-based schemes under this attack because these schemes change pseudonyms based on conditions that are external from the vehicle. Therefore, an adversary may try to mimic these conditions to force vehicles change pseudonyms frequently and deplete their pseudonyms pool. We simulate this attack by letting a random number of compromised vehicles drive within the road network. These vehicles act as LAA by changing their pseudonyms every 5 s and keep silent for 3 s and so on. This behavior is challenging the practicality of this attack because if the compromised vehicles

Table 7.5: CADS results under the LAA pseudonym depletion attack in sparse sub-dataset (silent neighbor threshold = 1; 3967 vehicles)

		LAA strength			
	No LAA	1%	3%	5%	10%
Compromised vehicles	0	40	119	198	397
Concerned vehicles or victims	2106	224	557	1041	1562
Average pseudonym lifetime (s)	114	88	85	80	74
Pseudonym change per Vehicle	1.3	1.8	1.8	1.8	1.9
Normalized distortion D_n (%)	44	56	59	58	58
QoS (%)	88	87	85	83	79

change their pseudonyms, they will suffer from self-depletion in short time when they use authenticated pseudonyms. If they use fake pseudonyms or do not change pseudonyms but switch to silence frequently, surrounding vehicles can detect this behavior and abandon the compromised vehicles from affecting their decisions. Regardless of the practicability issues, we assume here that the compromised vehicles own infinite number of authenticated pseudonyms and is able to change it freely.

In the worst case scenario, a victim vehicle will change its pseudonym every minimum pseudonym time, but the CADS and CAPS can reduce the effect of this attack through their parameter: the silent neighbor threshold. When the silent neighbor threshold is set to be more than one, the scheme requires several silent neighboring vehicles to switch to silence. This condition hinders the LAA attack since it is unlikely to have several LAA vehicles neighboring the victim vehicle. Also, CADS can employ the pseudonym time increment parameter to increase the minimum pseudonym time when the pseudonym is changed with a likely tracker confusion.

The CADS is evaluated against the LAA of different strengths in terms of the number of the compromised vehicles. The protection against this attack is measured by the number of pseudonym changes and the pseudonym lifetime made by vehicles on average. When calculating this metric, we considered only vehicles that met a LAA vehicle within 50 m radius for at least 15 s and changed their pseudonyms during simulation at least once. We selected the first and the last 5 min of the realistic traces and run simulation five times for each LAA strength with different compromised vehicles selected randomly. We selected 2 sub-datasets to show the effect of LAA on both sparse and dense traffic. These short traces will not affect the generality of the obtained results because we consider the pseudonym changing behavior rather than a full reconstruction

Table 7.6: CADS results under the LAA pseudonym depletion attack in dense sub-dataset (silent neighbor threshold = 2; 7390 vehicles)

		LAA strength			
	No LAA	1%	3%	5%	10%
Compromised vehicles	0	74	222	370	739
Concerned vehicles or victims	3526	744	2015	2946	3855
Average pseudonym lifetime (s)	156	142	132	122	103
Pseudonym change per Vehicle	1.1	1.2	1.2	1.3	1.4
Normalized distortion D_n (%)	43	47	49	49	53
QoS (%)	91	90	89	88	86

of long traces. We tested two thresholds of silent neighbors of 1 and 2 vehicles where all vehicles choose the normal privacy preference.

Table 7.5 shows the average metrics obtained using a silent neighbor threshold of one for the sparse sub-dataset. Four LAA strengths along with the case of no LAA are evaluated. The number of the compromised vehicles and the concerned vehicles, on which the given metrics are calculated, are listed in the first two rows of Table 7.5. The concerned vehicles are those changed their pseudonyms at least once and refer to the victim vehicles when LAA is present or all vehicles for the no LAA case. The next two rows show the average pseudonym lifetime and the number of pseudonyms changed per vehicle. It can be observed that the victim vehicles changed pseudonyms 1.38 times more than the case of no LAA. This small increase in pseudonym changes cannot result in pseudonym depletion unless the LAA vehicles continuously follow the victim vehicles. Furthermore, we show the distortion and QoS metrics for each case. Interestingly, the normalized distortion metric D_n is increased when the LAA is present because the compromised vehicles force surrounding vehicles to change pseudonyms. The increased pseudonym changes result in a decrease in QoS depending on the LAA strength. We repeated this experiment with a silent neighbor threshold of 2 but we found that the distortion is significantly reduced because it is rarely to find two silent neighbors in this sparse traffic.

Table 7.6 shows the average metrics obtained using a silent neighbor threshold of 2 for the dense sub-dataset. We use here a threshold of 2 because the traffic is dense and it is common to meet with a compromised vehicle repeatedly. We observe that the victim vehicles changed pseudonyms 1.27 times more than the case of no LAA at maximum. The same behavior of increased distortion and slight reduction in QoS is also observed.

From these observations, we conclude that a weak LAA of small percent of compromised vehicles (e.g., up to 3%) does not add a significant risk of pseudonyms depletion specially when setting the silent neighbor threshold to more than one. Also, this attack may hinder the threat of the GPA attack with a small impact on the QoS of safety applications.

7.5.5 CADS Efficiency

Regarding the efficiency of the CADS, we used the same Intel QuadCore i7-4800MQ @ 2.70GHz Hyper-threaded CPU and calculated the average running time of processing a single time step for one vehicle as we did with CAPS. We found that it takes 5.5 ms on average, which is again computationally efficient when a high-end CPU is used in the vehicle. However, if lower-end CPUs are used, then further code optimization should be investigated.

7.6 Comparative Evaluation

In this section, selected privacy schemes are evaluated and compared with our context-based schemes. We first evaluate SLOW [28], CSP [130] and CPN [97] quantitatively and evaluate mix zones qualitatively. We then compare these schemes along with our schemes in Section 7.6.4. A comparative evaluation with the RSP scheme [71] is already presented in Sections 7.4.3 and 7.5.4. Also, the tracking vulnerability of the periodical pseudonym change is shown in Section 3.4.6 showing its ineffectiveness in preventing tracking.

7.6.1 SLOW Scheme

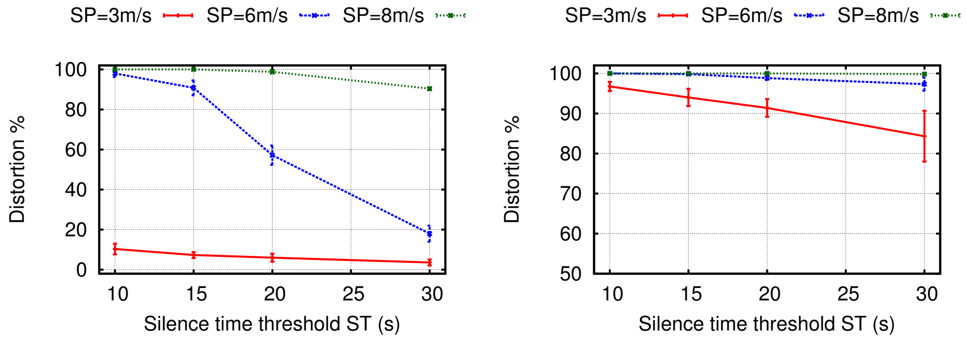
The pseudo code of the SLOW scheme is presented in Algorithm 4. In SLOW, a vehicle continuously checks its current speed and broadcasts beacons only when its speed is higher than a preset threshold SP. If a vehicle does not send beacons for ST time steps, it changes the pseudonym.

We evaluated the SLOW scheme in STRAW traces in both sparse and dense traffic, as shown in Figure 7.13. In sparse traffic of 50 vehicles, the distortion increases with the increase of the speed threshold (SP) because large thresholds let vehicles stop beaconing for long periods of time which, in turn, makes tracking difficult. The silent time threshold (ST) is relevant with the intermediate SP of 6 m/s because relatively short ST ($\leq 15s$) makes vehicles change pseudonyms frequently which, in turn, increases tracker confusion. In large ST, pseudonyms are only changed every long period which increases tracking and reduces the distortion. In dense traffic of 200 vehicles, the distortion is further increased because of the expected low speeds in dense traffic. All

Algorithm 4 SLOW scheme

```
Input: $SP, ST$
if $speed < SP$ then
    $silent\_time := silent\_time + 1$;
else
    if $silent\_time \geq ST$ then
        $psynm := get\_new\_pseudonym()$;
    end if
    $silent\_time := 0$;
    $send\_beacon()$;
end if
```

thresholds of SP and ST result in frequent pseudonym change and long silence which significantly increases tracking confusions. The success rates of tracking presented in [28] are much lower than ours because of the simplicity of their tracker model. Their attacker uses information of the last two beacons to calculate the acceleration of the vehicles.

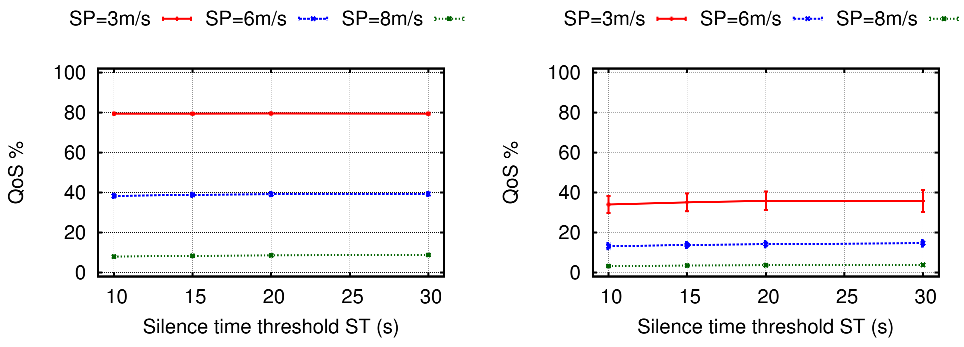


(a) Sparse: 50 vehicles

(b) Dense: 200 vehicles

Figure 7.13: Privacy level of SLOW in STRAW traces.

The QoS of SLOW is significantly reduced especially with relatively large speed thresholds (SP) ($>$ 3 m/s) in sparse traffic and with all thresholds in dense traffic, as shown in Figure 7.14. This significant reduction occurs because of the large amount of eliminated beacons during low speeds. Buttyán *et al.* [28] claimed that keeping silent at low speeds is safe because crashes occurring at low speeds cause fewer fatalities. However, turning off the transmitter reduces the awareness of other (fast) vehicles about slower vehicles which challenges safety applications.



(a) Sparse: 50 vehicles

(b) Dense: 200 vehicles

Figure 7.14: QoS level of SLOW in STRAW traces.

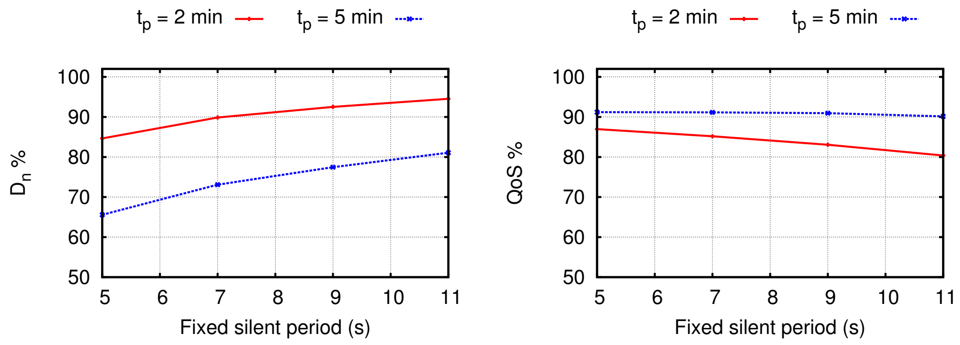
7.6.2 CSP Scheme

Coordinated Silent Period (CSP) is proposed by Tomandl *et al.* [130] in their comparison of silent period and mix zone schemes. CSP coordinates all vehicles in the network to remain silent and change pseudonyms synchronously. CSP seems to be theoretical since the coordination overhead in real world situations increases dramatically [130]. However, CSP increases the privacy significantly because it maximizes the size of the anonymity set at every pseudonym change.

In Figure 7.15, CSP is evaluated in realistic traces using two pseudonym lifetimes (t_p): 2 and 5 min. The normalized distortion (D_n) increases as the silent period increases because longer silence periods give a sufficient time for vehicles to change their states from those predicted by a tracker which, in turn, increases tracker confusions. Also, D_n increases as the lifetime of the pseudonym (t_p) decreases because shorter lifetimes increase the frequency of changing pseudonyms, and thus, tracker confusion. The QoS when t_p equals 5 min is almost constant because the silent periods are repeated only 6 times for the whole simulation resulting in fewer incorrect estimations by the in-vehicle tracker. When t_p equals 2 min, the QoS slightly decreases as the silent period increases.

7.6.3 CPN Scheme

The pseudo code of the Cooperative Pseudonym change scheme based on the number of Neighbors (CPN) [97] is presented in Algorithm 5. In CPN, vehicles monitor their neighbors within radius R and wait until they reach a threshold K. When this trigger occurs, the vehicle sets an internal flag $ready\_flag$, broadcasts this flag within the beacon and changes the pseudonym in the next



(a) Privacy

(b) QoS

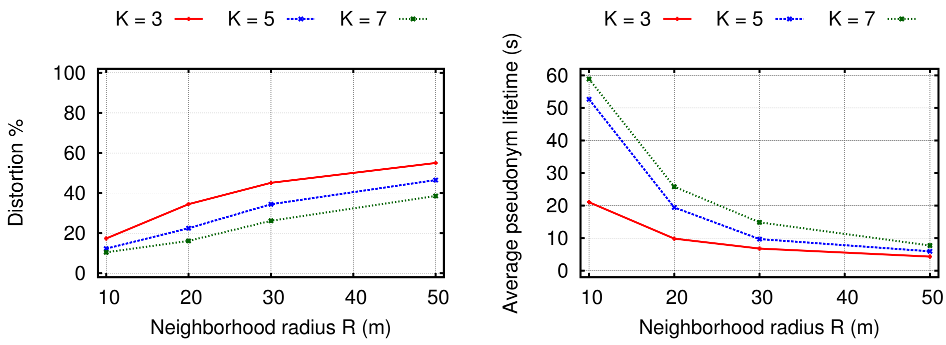
Figure 7.15: Privacy and QoS levels of CSP in realistic traces.

beacon. When a vehicle receives a beacon with a set flag or its internal flag is set already, it changes pseudonym immediately.

Algorithm 5 CPN Scheme

```
Input: $scan, ready\_flag, R, K$
$neighbors := get\_neighbors(scan, R);$
if $ready\_flag == 1$ then
    $psynm := get\_new\_pseudonym();$
    $ready\_flag := 0;$
else if $IsAnyNeighborReady(neighbors) == True$ then
    $psynm := get\_new\_pseudonym();$
else if $SIZE(neighbors) \geq K$ then
    $ready\_flag := 1;$
end if
$send\_beacon(psynm, ready\_flag);$
```

The distortion of CPN in the realistic traces is presented in Figure 7.16(a). The distortion increases as the neighborhood radius increases because in large radii, a vehicle could find more neighbors whose ready flag is set which, in turn, makes a vehicle change pseudonyms frequently. Distortion also increases as the threshold of number of neighbors (K) decreases because small thresholds make vehicles trigger to change pseudonyms more frequently. It is important to note that higher distortion levels of CPN are achieved through frequent pseudonym changes. In Figure 7.16(b), the average pseudonym lifetime versus neighborhood thresholds is depicted. We notice that the pseudonym lifetime decreases exponentially to achieve almost linear distortion levels as



(a) Privacy

(b) Pseudonym lifetime

Figure 7.16: (a) Privacy level and (b) Average pseudonym lifetime of CPN in realistic traces.

shown in Figures 7.16(a) and 7.16(b). For example, when a distortion level of 55% is achieved, vehicles need to change their pseudonym every 4 s on average, which requires a very large number of pseudonyms to be loaded in vehicles and maintained in the central authority.

The QoS of CPN is presented in Figure 7.17. The QoS is almost constant (around 91%) for all threshold values because the in-vehicle tracker is able to estimate the state of nearby vehicles very well. This high quality of estimation is the result of beacons being broadcast at every time step. Even when confusion occurs and the in-vehicle tracker mixes beacons, the confusion usually with a vehicle in close proximity whose state is similar to that of the correct vehicle. Therefore, the error in the estimated states are usually small which is reflected in a high QoS.

7.6.4 Comparison

In this section, we provide a quantitative comparison between different privacy schemes and our context-based schemes. Based on experiments performed on realistic traces, we made the following steps to align and compare the performance of privacy schemes. We rounded the QoS to the nearest integer. Then, the maximum (normalized) distortion that can be achieved in each QoS level is selected along with the average pseudonym lifetime performed by vehicles to achieve this maximum distortion. Figure 7.18 illustrates this comparison among CPN, CSP and RSP schemes along with our context-based schemes CAPS and CADS. The SLOW scheme is left out because it results in low QoS levels, as shown in Figure 7.14. The average pseudonym lifetime in seconds is

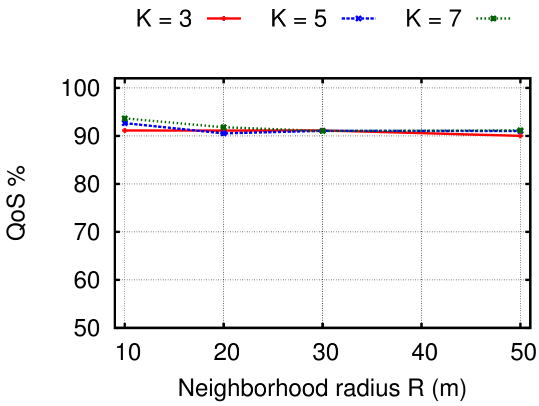
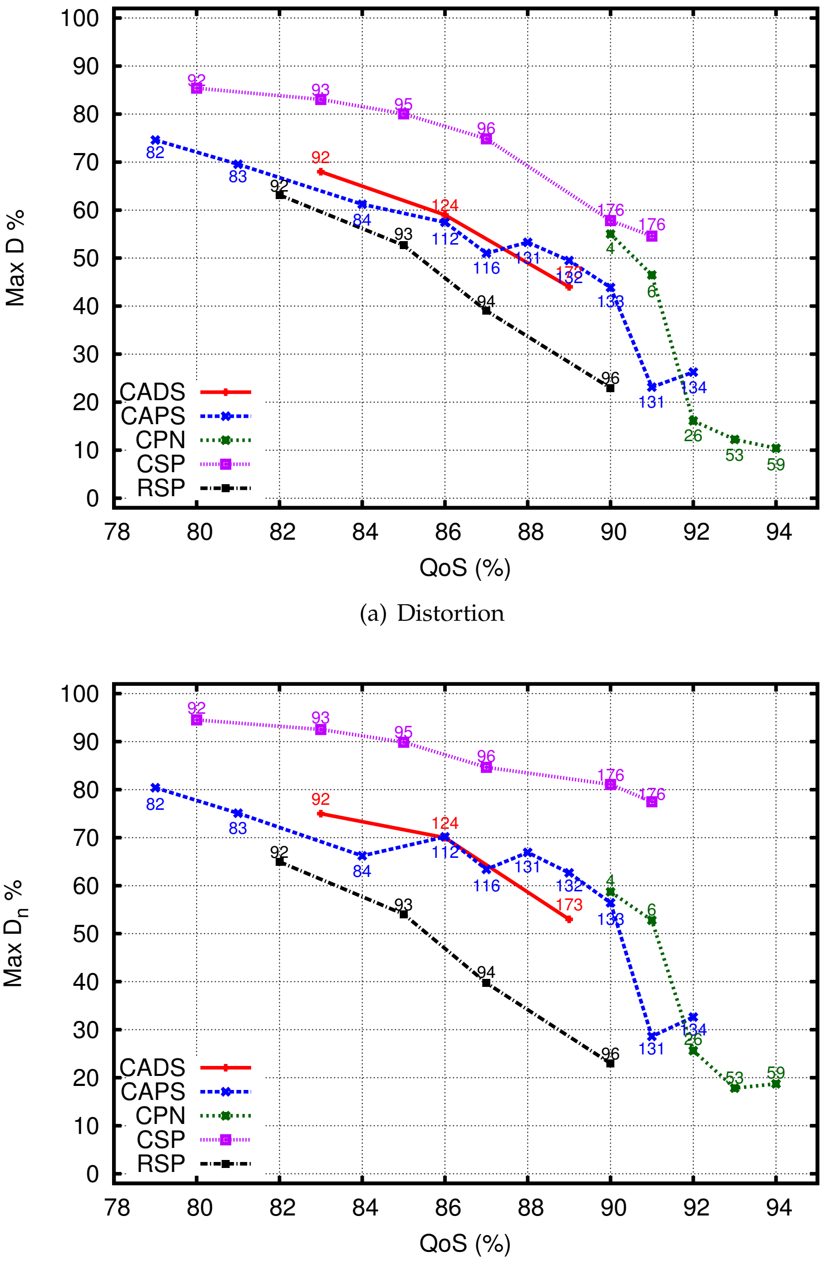


Figure 7.17: QoS level of CPN in realistic traces.

written over or under the graph lines.

The CSP provides the highest distortion among all other schemes given a similar QoS level. It results in a high QoS of up to 91% and requires a reasonable average pseudonym lifetime of about 3 min to achieve these high distortion and QoS levels. However, a global time synchronization among vehicles is challenging. Also, further investigation is required to study possible implications or attacks of this global synchronized silence. The delivery of packets and handling safety-critical situations during the scheduled silence are just examples that make the CSP unpractical. The next scheme is the CPN which results in the highest QoS levels over all other schemes (because it does not employ any silence before a pseudonym change). It can result in high distortion levels but with a significantly short pseudonym lifetime of 4 s. This is a serious drawback of CPN because it requires so frequent pseudonym changes to preserve privacy. It requires approximately 657,000 keys per year to be loaded in each vehicle (assuming 2 driving hours per day). This huge number of keys cannot be affordable by the certification authority which makes CPN impractical as well. The RSP achieves a good distortion level but with the cost on the QoS. Higher QoS levels can be attained but with low distortion levels.

The CADS and CAPS provide practical compromises among the distortion, QoS and average pseudonym lifetime. The performance of CAPS varies according to the provided parameters. CAPS can provide about 60% of normalized distortion when the QoS is 90%. The average pseudonym lifetime ranges from 1.3 min to 2.2 min depending on the achieved distortion and QoS levels. CADS gives the choice to drivers which privacy level matches with their preferences. An intermediate privacy preference results in distortion of 60% and QoS of 86%. The average pseudonym life time ranges from 1.5 min to 3 min.



(a) Distortion

(b) Normalized Distortion

Figure 7.18: Distortion versus QoS levels of different VANET privacy schemes in realistic traces. The average pseudonym lifetime for the maximum (normalized) distortion is written in seconds.

7.6.5 Mix Zone

We evaluate mix zones qualitatively because they are usually evaluated against timing and transition attacks. Since our tracker does not support these attacks, quantitative evaluation will not represent the actual performance of these schemes.

Mix zones are usually placed at road intersections since vehicle movements are not predictable. Within a mix zone, the exchanged beacon messages must be encrypted [53], or vehicles must be silent [27]. If vehicles change their pseudonyms within the mix zone, the adversary cannot correlate leaving vehicles to those entering the zone earlier because movement cannot be predicted. Mix zones have the following drawbacks if compared to our proposed context-based schemes:

- **Vulnerability to timing and transition attacks.** Since mix zones are placed in fixed locations, they are vulnerable to timing and transition attacks. An adversary can utilize additional knowledge about the timing and transition among different entry and exit points of the intersection. This knowledge can be obtained by visually monitoring the intersection and constructing a joint probability distribution for transition and timing. Using this distribution, the adversary can guess the mapping between the entering and leaving vehicles and thus correlate old and new pseudonyms. These attacks are effective. For example, Buttyán *et al.* [27] showed that a tracking success rate of up to 70% can be achieved by covering only half of intersections. In addition to timing and transition attacks, statistical features of vehicles, such as the driving behavior and the average speed before and after mix zones, can be employed to identify vehicles as shown in [151]. Our proposed schemes are not vulnerable to these attacks because silent periods are established dynamically based on the vehicle context and can happen in any part of the road.

- **RSU dependability.** Mix zones depend on RSUs to coordinate silence period or distribute encryption keys. However, it is not expected that RSUs will be widespread deployed especially in the initial deployment of VANET. CAPS and CADS let vehicles decide autonomously with no need for RSUs when and where a pseudonym should be changed.

- **Vulnerability to active attacks for cryptographic zones.** An active attacker may participate in the cryptographic mix zones and obtain the shared key. Once the key is obtained, the mix zone becomes useless because all exchanged messages can be observed and decrypted by a global adversary. Regarding our proposed schemes, an active adversary must compromise many vehicles (more than 1% of the vehicles) to be able to affect the pseudonym change frequency, as shown in Section 7.5.4. Also, the behavior of the active adversary works in the opposite interest of the global adversary because it forces vehicles to change pseudonyms more frequently which hinders the tracking attack. Thus, in our schemes, active and passive attacks cannot collude to track vehicles.

- **Safety concerns for silence-based zones.** Road intersections or joints are risky places in the road networks. In fact, intersection crashes represent 26% of all crashes [64]. Silence-based mix zones challenges this fact because it is inappropriate to remain silence in places where it is important to exchange safety messages.

7.7 Summary

In this chapter, we proposed two context-based location privacy schemes (CAPS and CADS) that significantly increase the distortion in both STRAW and realistic traces. They utilize a context monitoring module to track surrounding vehicles and identify adequate situations to change pseudonym and determine the effective length of silence period. In CADS, a driver can choose the desired privacy level and the scheme can automatically identify the appropriate parameters that fit this desired level based on the real-time traffic density. Based on the experiment results, CADS can increase distortion compared with the CAPS when normal or high privacy levels are selected with a slight reduction in the QoS. Also, the CADS can preserve highest distortion for vehicles that select a high privacy level even when they drive within a majority of vehicles selected a lower privacy level. Based on these results, choosing the appropriate context for changing pseudonyms is crucial to achieve high levels of both privacy and safety. Last but not least, various privacy schemes are evaluated and compared with our context-based schemes. CAPS and CADS showed a practical and reasonable compromise among privacy, QoS and the average pseudonym lifetime.

8 Conclusion and Future Work

8.1 Findings and Limitations

In this dissertation, we investigated location privacy in VANET and considered the impact of privacy schemes on the QoS of safety applications. VANET will be realized in the near future due to its numerous benefits to traffic safety and efficiency. Privacy of drivers must be well-protected to ensure the public acceptance of VANET. Despite there are some privacy schemes that are published in the literature, the impact of privacy schemes on safety applications is overlooked and sporadically measured by generalized network or error-based metrics. Also, there is no consensus on the privacy metric and their calculation methods. Consequently, comparison among different privacy schemes in terms of well-developed privacy and safety metrics is missing. We tried to fill **these gaps** in this dissertation. We worked toward our objectives and investigated all research questions that are raised in Section 1.2 and we conclude here our findings and limitations of each question.

Objective O1: Robust Vehicle Tracker

In Chapter 3, we addressed the research questions **RQ1** and **RQ2** which consider the most efficient tracking algorithm for the VANET beaconing use case and the main factors that affect this tracking. According to related studies, location privacy is inversely proportional to the adversary capability of tracking vehicle movements. Tracking vehicles over a wide coverage of the road network and for long time facilitates the re-identification of the anonymous reconstructed traces; thus disclosing the drivers' places of interests and threatening their location privacy. Therefore, we developed a robust tracker that can be used in evaluating privacy schemes. We conclude the following:

- We developed a vehicle tracker based on the NNPDA algorithm that uses the pseudonyms and the spatiotemporal information in beacon messages to reconstruct actual vehicle traces.

- This tracker is evaluated with different vehicle traces of various densities, position noises, beaconing rates and packet delivery ratios.

- The experiment results show that anonymous beacons sent every 1 s with position noise up to 1 m can be effectively tracked regardless of the vehicle density. Anonymous beaconing is equivalent to using a new pseudonym in every beacon which represents the most frequent pseudonym change possible.

- Pseudonymous beacons, where pseudonyms are changed every a period of time, are more accurately traceable even with large position noises up to 10 m.

- Our tracker achieved a higher accuracy than the MHT tracker, that is commonly-used in related work, in both noiseless and noisy positions.

- The position and velocity are the sufficient and necessary information to effectively track anonymous beacon messages.

- The main factors that reduce beacons traceability are the shorter pseudonym lifetime, higher vehicle density, less precise positions (noise $\geq$ 2 m), packet losses (PDR $\leq$ 80%) and lower beaconing rate ($<$ 1 Hz).

These results lead to the following findings:

- Since safety applications require a beaconing rate of up to 10 Hz, a position noise up to 1 m and authenticated beacons with certified pseudonyms, all these requirements facilitate continuous and accurate vehicle tracking. This clearly highlights the trade-off between the safety application requirements and location privacy and strongly supports the relevance of this dissertation.

- Simultaneous pseudonym changes among nearby vehicles do not necessarily cause tracker confusion because the spatiotemporal information can be employed to correlate old and new pseudonyms.

- Consequently, frequent pseudonym changes do not guarantee a better location privacy since the tracker is not confused at every change.

- Anonymity set should not be defined as the nearby vehicles that change their pseudonyms simultaneously. This definition is misleading and overestimates the gained privacy because the tracker can effectively discriminate between members of this set.

- The high accuracy of our tracker confirms its capability and suitability to act as a global adversary for location privacy evaluation.

- The high efficiency of our tracker lets us propose embedding it inside vehicles which enhances the vehicle awareness about its surrounding traffic and help in evaluation of the likelihood of tracker confusion.

However, these findings are restricted by the following limitations:

- Although the high accuracy of the developed tracker, it can be further enhanced. The tracker considers only the pseudonym and the spatiotemporal information included in beacons. There are other important beacon information that can be exploited to discriminate among mixed beacons such as the vehicle type and size. Also, the road map and geometry can be used to better predict the vehicle state especially after silence periods. However, these enhancements will reduce the efficiency of tracker.
- When using this tracker as a global observer, it will not so effective in evaluating mix zone privacy schemes. Our tracker does not include timing and transition attacks that can be posed at road intersections. These attacks are essential to measure the effectiveness of the mix zone schemes.

Objective O2: Suitable Privacy Metric

The research question **RQ3** considers measuring location privacy. In Chapter 4, we investigated different location privacy metrics used in VANET domain. In fact, each metric is calculated differently in different research works and evaluated using different adversary models. We reviewed typical metrics and conclude the following:

- Four location privacy metrics are discussed and reviewed which are anonymity set size, entropy, traceability and distortion.
- Traceability and distortion metrics are thoroughly investigated and formally defined to reflect the best knowledge that the adversary can obtain to re-identify the reconstructed traces.
- To compare these metrics, we employed the random silent period scheme with three parameter sets which expectedly result in low, intermediate and high privacy levels, respectively. We used our tracker to reconstruct vehicle traces from beacons altered by this privacy scheme. We then measured these four metrics for each parameters set.

According to the comparison results, we found the following:

- The anonymity set size is unsuitable in measuring location privacy because it does not show any variation with different strengths of a privacy scheme. The entropy is also not a good candidate because it does not provide a unified variation in different traffic densities.

- Traceability and distortion are appropriate metrics, but the distortion metric filters out traces that are completely tracked but not similar to the original traces. We assume that the more similar the reconstructed traces to the original ones, the more successful the tracker in re-identifying these pseudonymous traces and threatening the drivers' privacy.

- Employing the distortion metric in measuring location privacy increases the trustworthiness in the results presented in this dissertation when compared with research works that use unsuitable metrics.

However, the proposed distortion metric is restricted by the following limitations:

- This metric is calculated based on the output of our tracker. Other advanced trackers may result in lower distortion levels. This means the presented evaluations of privacy schemes represent an upper-bound rather than lower-bound location privacy.

- The proposed metric assumes a global adversary who seeks to reconstruct vehicle traces as accurate and complete as possible to be able to re-identify them effectively. This metric does not measure location privacy against other adversaries that have different objectives or exploit knowledge from other sources. These adversaries may fulfill their objectives even with high distortion levels depending on the type of the attack.

Objective O3: Impact on Safety Applications

The research question **RQ4** considers measuring the impact of privacy schemes on safety applications. Privacy schemes usually eliminate beacons during silence periods which reduces the awareness of the vehicle about the surrounding traffic which, in turn, decreases the effectiveness of safety applications significantly. Despite the importance of measuring this impact, it is rarely considered in the literature. Therefore, we thoroughly investigated this issue in Chapter 5 and conclude the following:

- Two safety applications are considered which are forward collision warning and lane change warning applications because they require the most precise information and the most frequent beaconing rate.

- We proposed a generic methodology that measures the quality of service (QoS) by calculating the probability of correctly estimating the requirements of a safety application. To calculate this probability, this methodology uses the expected errors of beacon information after applying a privacy scheme in a Monte Carlo analysis.
- To estimate the error in beacon information, we assume that a local tracker is embedded inside vehicles that monitors the nearby vehicles through their broadcast beacons. This in-vehicle tracker accurately estimates the states of surrounding vehicles even when position noises are present or their beacons are missed due to a network error or a silence period.

According to the experiment results, we found the following:

- The proposed methodology is generally applicable to any privacy scheme because Monte Carlo calculations work directly on error samples obtained from tracking of beacons modified by a privacy scheme.
- This methodology is also extensible to other applications provided that the application requirements can be formulated as equations in terms of error samples of vehicle states.
- Using a local tracker inside vehicles relaxes the requirements of safety applications. A reasonably high QoS can be achieved even with lower beaconing rates and imprecise position information.

However, the proposed QoS measurement methodology is restricted by the following limitations:

- The proposed QoS metric does not measure the quality and timing of the alerts of a safety application. We are not certain about the effect of different QoS levels on providing timely and correct alerts because it depends on the design of the application to a large extent.
- This metric describes the general performance of the whole VANET scenario rather than specific critical situations. It cannot provide the expected performance of individual vehicles.

Objective O4: Propose Privacy Schemes

We proposed several privacy schemes in this dissertation. In Chapter 6, we investigated and proposed **obfuscation** schemes which address some conclusions and findings for the research question **RQ5** as follows:

- The proposed obfuscation schemes add large position noises for a random short period after a pseudonym change and broadcast beacons over a random rate.
- Large noises added after a pseudonym change can be skipped by the adversary by ignoring beacons of a new pseudonym for a while. No noise is added to the vehicle state when the same pseudonym is used because this noise causes no tracker confusion and can be easily filtered. Thus, information perturbation schemes are ineffective in preserving location privacy in VANET.
- Random beaconing rates are also ineffective in preserving privacy because beacons of consecutive time steps can be merged together to formulate beacons of all vehicles over longer time steps. The merged beacons can then be tracked effectively. Besides, these random rates reduce the QoS of safety applications significantly because they eliminate large number of beacons every time step.

We addressed the research question **RQ6** which considers context-based privacy schemes in Chapter 7. We proposed two schemes that choose the appropriate context to remain silent and change pseudonyms so that the likelihood of tracker confusion is increased. We conclude the following:

- These schemes use an in-vehicle tracker to provide a more realistic view about the surrounding traffic and facilitate estimating the likelihood of tracker confusion.
- The context-aware privacy scheme (CAPS) allows a vehicle to select the effective context in which a vehicle should remain silent and change its pseudonym and when to resume beaconing with a high probability of confusion to a global adversary.
- CAPS was further enhanced by proposing the context-adaptive scheme (CADS) which selects an optimized parameters set for CAPS based on the real-time traffic density and user privacy preference. CADS can keep a high distortion level for vehicles that select a high privacy preference even when they drive within a majority of vehicles selected a lower privacy preference.

Based on these conclusions, we found the following

- Choosing the appropriate context to change pseudonyms and remaining silent for a sufficient period are two essential factors to increase the likelihood of tracker confusion. They avoid useless pseudonym changes and unnecessary long silent periods which, in turn, results in a higher QoS of safety applications.

- Privacy consideration and recognition differ from person to another and it is beneficial to employ this fact to relax some privacy restrictions to enhance the safety of the whole system.

However, the proposed context-based schemes are restricted by the following limitations:

- Although they showed an efficient performance on the development machine, they need to be tested on hardware testbed with specifications that are expected in an automotive environment.
- The distortion achieved by the context-based schemes is not considerably high. However, we are not certain about how successful the attacks posed with this level of distortion.

Objective O5: Privacy Schemes Comparison

The last but not least objective and research question **RQ7** consider evaluation of the existing privacy schemes. In Section 2.6, we provided a thorough review of different approaches of privacy schemes. In Section 7.6, we provided quantitative and qualitative evaluations for privacy schemes showing their privacy and QoS levels. Based on these evaluations, we conclude the following:

- Coordinated silent period scheme provides high distortion and QoS levels by remaining silent synchronously and globally among all vehicles before a pseudonym change. However, a global coordination among vehicles is challenging and needs further investigation regarding possible attacks or implications of this global synchronized silence.
- Cooperative pseudonym change scheme can result in a good distortion level with a reasonably high QoS but with very short pseudonym lifetime which makes it impractical.
- Both CAPS and CADS provide a more practical compromise among acceptable distortion and QoS levels and relatively long pseudonym lifetime.
- Although the effectiveness of mix zones in reducing beacons traceability, they suffer from some issues such as transition and timing attacks, active attacks and dependability on road-side units.

8.2 Future Work

The results of this thesis, with contributions and limitations, indicate that it is possible to design schemes that effectively preserve privacy with a minimal impact on safety applications. In this direction, some possible future works can be pursued, as described next:

- **Incorporate safety conditions inside context-based schemes.** Since vehicles will include a tracker that can effectively monitor and track nearby vehicles, it would be a good advancement for vehicles to identify safety-critical situations and stop privacy-preserving operations in these situations. Also, non-critical situations should be recognized to allow privacy schemes to operate freely. This research direction opens several new challenges. For example, if the privacy is dynamically controlled by external conditions, how to secure vehicles from bogus attacks that try to prevent vehicles from enabling privacy schemes? Also, what is the safety level that should be considered as critical? And who should determine that threshold? Should the recognition of safety level be adaptive according to the road conditions, drivers' experience or vehicle capabilities?

- **Integrate several privacy models into one general privacy protocol.** As discussed and evaluated in this dissertation, no privacy scheme has an absolute advantage over all others especially in handling the trade-off between privacy and safety. It would be valuable to integrate different schemes in a single large scenario to take the advantage of all schemes. For example, deploying cryptographic mix-zones in the city center where RSUs may be widespread installed. In other regions, vehicles should enable context-based schemes where no RSU is available. Also, vehicles should cooperate to establish a local cryptographic group while they drive in highways where the network topology is somehow stable.

- **Deployment Issues.** There are some open issues regarding deployment of privacy schemes in real-world scenarios. First, context-based privacy schemes should be evaluated on automotive testbeds to study computation and communication limitations. Second, how should privacy schemes handle low penetration rate scenarios that are expected in the initial deployment phase of VANET? Third, privacy schemes should be also evaluated against a weaker but practical adversary who covers only some parts of the road network. Especially, what kind of attacks can this adversary perform against vehicles? How can the privacy be measured in this case, provided that complete traces cannot be reconstructed?

Bibliography

[1] Connected Vehicle Safety Pilot. http://www.its.dot.gov/safety_pilot/. [Online; accessed Sep-2015].

[2] Drive C2X. http://www.drive-c2x.eu/. [Online; accessed Sep-2015].

[3] Safe and Intelligent Mobility - Test Field Germany (simTD). http://www.simtd.de/. [Online; accessed Sep-2015].

[4] Safe Road Trains for the Environment (SARTRE). http://www.sartre-project.eu/. [Online; accessed Sep-2015].

[5] ETSI TR 102 638 V1.1.1. *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions*, Jun 2009.

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

[7] TAPASCologne project, 2010. [Online; accessed 20-January-2015].

[8] ETSI TS 102 867 v1.1.1. *Intelligent Transport Systems (ITS); Security; Stage 3 mapping for IEEE 1609.2*, Jun 2012.

[9] ETSI TS 102 940 V1.1.1. *Intelligent Transport Systems (ITS); Security; ITS communications security architecture and security management*, Jun 2012.

[10] ETSI TS 102 941 V1.1.1. *Intelligent Transport Systems (ITS); Security; Trust and Privacy Management*, Jun 2012.

[11] Ieee standard for wireless access in vehicular environments security services for applications and management messages. *IEEE Std 1609.2-2013 (Revision of IEEE Std 1609.2-2006)*, pages 1–289, April 2013.

[12] DSRC: The Future of Safer Driving. Fact Sheet. http://www.its.dot.gov/factsheets/dsrc_factsheet.htm, Sept 2015. [Online; accessed Oct-2015].

[13] S Al-Sultan and MM Al-Doori. A comprehensive survey on vehicular Ad Hoc network. *Journal of Network and Computer Applications*, 37:380–392, 2014.

[14] Nikolaos Alexiou, Marcello Laganà, Stylianos Gisdakis, Mohammad Khodaei, and Panagiotis Papadimitratos. Vespa: Vehicular security and privacy-preserving architecture. In *Proceedings of the 2nd ACM workshop on Hot topics on wireless network security and privacy*, pages 19–24. ACM, 2013.

[15] Gianmarco Baldini, Vincent Mahieu, Igor Nai Fovino, Alberto Trombetta, and Marco Taddeo. Identity-based security systems for vehicular ad-hoc networks. In *Connected Vehicles and Expo (ICCVE), 2013 International Conference on*, pages 672–678. IEEE, 2013.

[16] Y. Bar-Shalom, F. Daum, and J. Huang. The probabilistic data association filter. *Control Systems, IEEE*, 29(6):82 –100, December 2009.

[17] Thomas Benz, Antonio Kung, Martin Kost, Frank Kargl, Zhendong Ma, Guido Tijskens, and J.C. Freytag. Preciosa: V2x privacy issue analysis, 2009. Deliverable 1. Available from `http://www.transport-research.info/Upload/Documents/201210/20121025_103828_50034_PRECIOSA_D1_V2XPrivacyIssuesAnalysis_v4.1.pdf` [accessed Oct 2015].

[18] Alastair R Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive computing*, 2(1):46–55, 2003.

[19] A.R. Beresford and F. Stajano. Mix zones: user privacy in location-aware services. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 127–131, March 2004.

[20] Carl Bergenhem, Steven Shladover, Erik Coelingh, Christoffer Englund, and Sadayuki Tsugawa. Overview of platooning systems. In *Proceedings of the 19th ITS World Congress, Oct 22-26, Vienna, Austria (2012)*, 2012.

[21] Laurent Bindschaedler, Murtuza Jadliwala, Igor Bilogrevic, Imad Aad, Philip Ginzboorg, Valtteri Niemi, and Jean-Pierre Hubaux. Track me if you can: On the effectiveness of context-based identifier changes in deployed mobile networks. In *NDSS*, 2012.

[22] Subir Biswas, Jelena Mišić, and Vojislav Mišić. An identity-based authentication scheme for safety messages in wave-enabled vanets. *International Journal of Parallel, Emergent and Distributed Systems*, 27(6):541–562, 2012.

[23] S.S. Blackman and R. Popoli. *Design and analysis of modern tracking systems*. Artech House radar library. Artech House, August 1999.

[24] Dan Boneh and Matt Franklin. Identity-based encryption from the weil pairing. In *Advances in Cryptology-CRYPTO 2001*, pages 213–229. Springer, 2001.

[25] R Braun, F Busch, C KEMPER, R HILDEBRANDT, F WEICHENMEIER, C MENIG, I PAULUS, and R PRESSLEIN-LEHLE. Travolution - netzweite optimierung der lichtsignalsteuerung und lsa-fahrzeugkommunikation (in english: Travolution - network-wide optimization of traffic signal control and traffic signal to vehicle communication). In *Strassenverkehrstechnik*, volume 53, pages 365–74. Forschungsgesellschaft fuer Strassen- und Verkehrswesen (FGSV), 2009.

[26] M. Burmester, E. Magkos, and V. Chrissikopoulos. Strengthening privacy protection in vanets. In *Networking and Communications, 2008. WIMOB '08. IEEE International Conference on Wireless and Mobile Computing,*, pages 508–513, Oct 2008.

[27] Levente Buttyán, Tamás Holczer, and István Vajda. On the effectiveness of changing pseudonyms to provide location privacy in vanets. In *Proceedings of the 4th European Conference on Security and Privacy in Ad-hoc and Sensor Networks*, ESAS'07, pages 129–141, Berlin, Heidelberg, 2007. Springer-Verlag.

[28] Levente Buttyán, Tamas Holczer, Andre Weimerskirch, and William Whyte. SLOW: A Practical pseudonym changing scheme for location privacy in VANETs. In *2009 IEEE Vehicular Networking Conference (VNC)*, pages 1–8. IEEE, October 2009.

[29] Levente Buttyán and Jean-Pierre Hubaux. *Security and Cooperation in Wireless Networks: Thwarting Malicious and Selfish Behavior in the Age of Ubiquitous Computing*. Cambridge University Press, 2007.

[30] Giorgio Calandriello, Panos Papadimitratos, Jean-Pierre Hubaux, and Antonio Lioy. Efficient and robust pseudonymous authentication in VANET. In *Proceedings of the fourth ACM international workshop on Vehicular ad hoc networks - VANET '07*, pages 19–28, New York, New York, USA, 2007. ACM Press.

[31] Giorgio Calandriello, Panos Papadimitratos, Jean-Pierre Hubaux, and Antonio Lioy. On the performance of secure vehicular communication

systems. *Dependable and Secure Computing, IEEE Transactions on*, 8(6):898–912, 2011.

[32] Derek Caveney. *Cooperative Vehicular Safety Applications*, pages 21–48. John Wiley and Sons, Ltd, 2009.

[33] Alket Cecaj, Marco Mamei, and Nicola Bicocchi. Re-identification of anonymized CDR datasets using social network data. In *The Third IEEE International Workshop on the Impact of Human Mobility in Pervasive Systems and Applications*, pages 237–242. Ieee, March 2014.

[34] David Chaum and Eugène Van Heyst. Group signatures. In *Advances in Cryptology-EUROCRYPT'91*, pages 257–265. Springer, 1991.

[35] David L Chaum. Untraceable electronic mail. *Return Addresses, and Digital Pseudonyms*, 24(2):84–90, 1981.

[36] David R Choffnes and Fabián E Bustamante. An integrated mobility and traffic model for vehicular wireless networks. In *Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks*, pages 69–78. ACM, Sept 2005.

[37] CAR 2 CAR Communication Consortium. Car 2 car communication consortium manifesto. Technical report, 2007.

[38] Vehicle Safety Communications Consortium. *Vehicle Safety Communications Project: Task 3 Final Report: Identify Intelligent Vehicle Safety Applications Enabled by DSRC.* National Highway Traffic Safety Administration, Office of Research and Development, Washington, D.C., 2005.

[39] Gabrielle Demange, David Gale, and Marilda Sotomayor. Multi-item auctions. *Journal of Political Economy*, 94(4):863–872, August 1986.

[40] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *Proceedings of the 2Nd International Conference on Privacy Enhancing Technologies*, PET'02, pages 54–68, Berlin, Heidelberg, 2003. Springer-Verlag.

[41] F Dressler, H Hartenstein, O Altintas, and O K Tonguz. Inter-vehicle communication: Quo vadis. *Communications Magazine, IEEE*, 52(6):170–177, 2014.

[42] D. Eckhoff, R. German, C. Sommer, F. Dressler, and T. Gansen. Slotswap: strong and affordable location privacy in intelligent transportation systems. *Communications Magazine, IEEE*, 49(11):126 –133, Nov. 2011.

[43] Tamer ElBatt, Siddhartha K. Goel, Gavin Holland, Hariharan Krishnan, and Jayendra Parikh. Cooperative collision warning using dedicated short range wireless communications. In *Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks*, VANET '06, pages 1–9, New York, NY, USA, 2006. ACM.

[44] Karim Emara, Wolfgang Woerndl, and Johann Schlichter. Beacon-based Vehicle Tracking in Vehicular Ad-hoc Networks. Technical report, TECHNISCHE UNIVERSITÄT MÜNCHEN, April 2013.

[45] Karim Emara, Wolfgang Woerndl, and Johann Schlichter. Vehicle tracking using vehicular network beacons. In *Fourth International Workshop on Data Security and PrivAcy in wireless Networks (D-SPAN)*, Madrid, Spain, June 2013.

[46] Karim Emara, Wolfgang Woerndl, and Johann Schlichter. CAPS: Context-Aware Privacy Scheme for VANET Safety Applications. In *Proceedings of the 8th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, WiSec '15, New York, NY, USA, 2015. ACM.

[47] Cristofer Englund, Lei Chen, Alexey Vinel, and ShihYang Lin. Future applications of vanets. In Claudia Campolo, Antonella Molinaro, and Riccardo Scopigno, editors, *Vehicular ad hoc Networks*, pages 525–544. Springer International Publishing, 2015.

[48] Richard Gilles Engoulou, Martine Bellaïche, Samuel Pierre, and Alejandro Quintero. Vanet security surveys. *Computer Communications*, 44:1–13, 2014.

[49] Lars Fischer, Stefan Katzenbeisser, and Claudia Eckert. Measuring unlinkability revisited. *Proceedings of the 7th ACM workshop on Privacy in the electronic society - WPES '08*, page 105, 2008.

[50] Robert J. Fitzgerald. Development of practical pda logic for multitarget tracking by microprocessor. In *American Control Conference*, pages 889–898, june 1986.

[51] Anthony Foxx. A Dialogue with Industry, a Conversation between Cars. `https://www.transportation.gov/fastlane/dialogue-industry-conversation-between-cars`, May 2015. [Online; accessed Sep-2015].

[52] Julien Freudiger, Mohammad Hossein Manshaei, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. On the Age of Pseudonyms in Mobile Ad Hoc

Networks. In *2010 Proceedings IEEE INFOCOM*, pages 1–9. Ieee, March 2010.

[53] Julien Freudiger, Maxim Raya, Márk Félegyházi, Panos Papadimitratos, and Jean-Pierre Hubaux. Mix-Zones for Location Privacy in Vehicular Networks. In *ACM Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS)*, Vancouver, August 2007.

[54] Julien Freudiger, Reza Shokri, and Jean-Pierre Hubaux. On the optimal placement of mix zones. In *Proceedings of the 9th International Symposium on Privacy Enhancing Technologies*, PETS '09, pages 216–234, Berlin, Heidelberg, 2009. Springer-Verlag.

[55] José María de Fuentes, Ana Isabel González-Tablas, and Arturo Ribagorda. Overview of security issues in vehicular ad-hoc networks. In *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts*, pages 894–911. IGI Global, 2010.

[56] R. Fukui, H. Koike, and H. Okada. Dynamic integrated transmission control (ditrac) over inter-vehicle communications in its. In *Vehicular Technology Conference, 2002. VTC Spring 2002. IEEE 55th*, volume 1, pages 483–487 vol.1, 2002.

[57] M. Gerlach and F. Guttler. Privacy in vanets using changing pseudonyms - ideal and real. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 2521–2525, April 2007.

[58] Saira Gillani, Farrukh Shahzad, Amir Qayyum, and Rashid Mehmood. A survey on security in vehicular ad hoc networks. In *Communication Technologies for Vehicles*, pages 59–74. Springer, 2013.

[59] Philippe Golle and Kurt Partridge. On the anonymity of home/work location pairs. In *Proceedings of the 7th International Conference on Pervasive Computing*, Pervasive '09, pages 390–397, Berlin, Heidelberg, May 2009. Springer-Verlag.

[60] PTV Group. Vissim 5.1. `http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/`, 2009.

[61] Marco Gruteser and Dirk Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *Proceedings of the 1st international conference on Mobile systems applications and services (MobiSys 03)*, pages 31–42, 2003.

[62] Jinhua Guo. Security and privacy in vehicular networks. In *National Workshop on High-Confience Automotive Cyber-Physical Systems*, 2008.

[63] Jinhua Guo, John P Baugh, and Shengquan Wang. A group signature based secure and privacy-preserving vehicular communication framework. *Mobile Networking for Vehicular Environments*, 2007:103–108, 2007.

[64] J. Harding, G. Powell, R. Yoon, J. Fikentscher, C. Doyle, D. Sade, M. Lukuc, J. Simons, and J. Wang. Vehicle-to-Vehicle Communications: Readiness of V2V Technology for Application. Technical report, National Highway Traffic Safety Administration, Washington, DC, August 2014.

[65] H. Hartenstein and K.P. Laberteaux. A tutorial survey on vehicular ad hoc networks. *Communications Magazine, IEEE*, 46(6):164 –171, jun 2008.

[66] B Hoh, M Gruteser, H Xiong, and A Alrabady. Enhancing Security and Privacy in Traffic-Monitoring Systems. *Pervasive Computing, IEEE*, 5(4):38–46, 2006.

[67] Baik Hoh and Marco Gruteser. Protecting location privacy through path confusion. In *Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks*, SECURECOMM '05, pages 194–205, Washington, DC, USA, 2005. IEEE Computer Society.

[68] Baik Hoh, Marco Gruteser, Ryan Herring, Jeff Ban, Daniel Work, Juan-Carlos Herrera, Alexandre M Bayen, Murali Annavaram, and Quinn Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *Proceeding of the 6th international conference on Mobile systems, applications, and services*, pages 15–28, 2008.

[69] Baik Hoh, Marco Gruteser, Hui Xiong, and Ansaf Alrabady. Preserving privacy in gps traces via uncertainty-aware path cloaking. In *Proceedings of the 14th ACM Conference on Computer and Communications Security*, CCS '07, pages 161–171, New York, NY, USA, 2007. ACM.

[70] Yih-Chun Hu and Kenneth P Laberteaux. Strong vanet security on a budget. In *Proceedings of Workshop on Embedded Security in Cars (ESCAR)*, volume 6, pages 1–9, 2006.

[71] Leping Huang, K. Matsuura, H. Yamane, and K. Sezaki. Enhancing wireless location privacy using silent period. In *Wireless Communications and Networking Conference, 2005 IEEE*, volume 2, pages 1187–1192 Vol. 2, March 2005.

[72] Leping Huang, Hiroshi Yamane, Kanta Matsuura, and Kaoru Sezaki. Silent cascade: Enhancing location privacy without communication qos degradation. In *Proceedings of the Third International Conference on Security in Pervasive Computing*, SPC'06, pages 165–180, Berlin, Heidelberg, 2006. Springer-Verlag.

[73] Leping Huang, Hiroshi Yamane, Kanta Matsuura, and Kaoru Sezaki. Towards modeling wireless location privacy. In George Danezis and David Martin, editors, *Privacy Enhancing Technologies*, volume 3856 of *Lecture Notes in Computer Science*, pages 59–77. Springer Berlin Heidelberg, 2006.

[74] J. Jakubiak and Y. Koucheryavy. State of the art and research challenges for vanets. In *Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE*, pages 912–916, Jan 2008.

[75] J.G. Jordan, F. Soriano, D. Graullera, and G. Martin. A comparison of different technologies for efc and other its applications. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, pages 1171–1176, Aug 2001.

[76] R.E. Kalman et al. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[77] G. Karagiannis, O. Altintas, E. Ekici, G. Heijenk, B. Jarupan, K. Lin, and T. Weil. Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions. *Communications Surveys Tutorials, IEEE*, 13(4):584 –616, nov 2011.

[78] M. Khodaei, Hongyu Jin, and P. Papadimitratos. Towards deploying a scalable amp; robust vehicular identity and credential management infrastructure. In *Vehicular Networking Conference (VNC), 2014 IEEE*, pages 33–40, Dec 2014.

[79] Chonlatee Khorakhun, Holger Busche, and Hermann Rohling. Congestion control for vanets based on power or rate adaptation. In *Proceedings of the 5th international workshop on intelligent transportation (WIT)*, 2008.

[80] Marie-Ange Lèbre, Frédéric Le Mouël, Eric Ménard, Julien Dillschneider, and Richard Denis. VANET applications: Hot use cases. *CoRR*, abs/1407.4088, aug 2014.

[81] Stephanie Lefevre, Jonathan Petit, Ruzena Bajcsy, Christian Laugier, and Frank Kargl. Impact of v2x privacy strategies on intersection collision avoidance systems. In *Vehicular Networking Conference (VNC), 2013 IEEE*, pages 71–78, Dec 2013.

[82] Fan Li and Yu Wang. Routing in vehicular ad hoc networks: A survey. *Vehicular Technology Magazine, IEEE*, 2(2):12 –22, jun 2007.

[83] Mingyan Li, Krishna Sampigethaya, Leping Huang, and Radha Poovendran. Swing & swap: user-centric approaches towards maximizing location privacy. In *Proceedings of the 5th ACM workshop on Privacy in electronic society*, pages 19–28, 2006.

[84] Yunxin(Jeff) Li. An overview of the dsrc/wave technology. In Xi Zhang and Daji Qiao, editors, *Quality, Reliability, Security and Robustness in Heterogeneous Networks*, volume 74 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 544–558. Springer Berlin Heidelberg, 2012.

[85] Jianxiong Liao and Jianqing Li. Effectively Changing Pseudonyms for Privacy Protection in VANETs. In *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks*, pages 648–652. Ieee, December 2009.

[86] Xiaodong Lin, Xiaoting Sun, Pin-Han Ho, and Xuemin Shen. Gsis: A secure and privacy-preserving protocol for vehicular communications. *Vehicular Technology, IEEE Transactions on*, 56(6):3442 –3456, November 2007.

[87] Rongxing Lu, Xiaodong Li, T.H. Luan, Xiaohui Liang, and Xuemin Shen. Pseudonym changing at social spots: An effective strategy for location privacy in vanets. *Vehicular Technology, IEEE Transactions on*, 61(1):86 –96, January 2012.

[88] Rongxing Lu, Xiaodong Lin, Haojin Zhu, Pin-Han Ho, and Xuemin Shen. Ecpp: Efficient conditional privacy preservation protocol for secure vehicular communications. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*. IEEE, 2008.

[89] Zhendong Ma, Frank Kargl, and Michael Weber. Measuring location privacy in V2X communication systems with accumulated information. *2009 IEEE 6th International Conference on Mobile Adhoc and Sensor Systems*, pages 322–331, oct 2009.

[90] JA Misener, Raja Sengupta, and H Krishnan. Cooperative collision warning: Enabling crash avoidance with wireless technology. In *12th World Congress on ITS*, pages 6–10, San Francisco, November 2005.

[91] Y. L. Morgan. Notes on DSRC & WAVE standards suite: Its architecture, design, and characteristics. *IEEE Communications Surveys and Tutorials*, 12(4):504–518, 2010.

[92] Wassim G. Najm, Jonathan Koopmann, John D. Smith, and John Brewer. Frequency of Target Crashes for IntelliDrive Safety Systems. Technical report, National Highway Traffic Safety Administration (NHTSA), Cambridge, MA, October 2010.

[93] National Highway Traffic Safety Administration (NHTSA). Advance notice of proposed rulemaking (ANPRM) (Docket No. NHTSA-2014-0022). `http://www.regulations.gov/#!documentDetail;D=NHTSA-2014-0022-0002`, August 2014. [Online; accessed Sep-2015].

[94] B. Palanisamy, S. Ravichandran, Ling Liu, Binh Han, Kisung Lee, and C. Pu. Road network mix-zones for anonymous location based services. In *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*, 2013.

[95] Balaji Palanisamy and Ling Liu. MobiMix: Protecting location privacy with mix-zones over road networks. In *IEEE 27th International Conference on Data Engineering*, pages 494–505. Ieee, April 2011.

[96] Balaji Palanisamy and Ling Liu. Attack-resilient Mix-zones over Road Networks: Architecture and Algorithms. *IEEE Transactions on Mobile Computing*, 14(3):495–508, 2015.

[97] Yuanyuan Pan and Jianqing Li. Cooperative pseudonym change scheme based on the number of neighbors in VANETs. *Journal of Network and Computer Applications*, 36(6):1599 – 1609, 2013.

[98] Yuanyuan Pan, Jianqing Li, Li Feng, and Ben Xu. An analytical model for random pseudonym change scheme in VANETs. *Cluster Computing*, 17(2):413–421, January 2013.

[99] Yuanyuan Pan, Jianqing Li, Li Feng, and Ben Xu. An analytical model for random pseudonym change scheme in vanets. *Cluster Computing*, 17(2):413–421, 2014.

[100] P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya, Zhendong Ma, F. Kargl, A. Kung, and J.-P. Hubaux. Secure vehicular communication systems: design and architecture. *Communications Magazine, IEEE*, 46(11):100 –109, November 2008.

[101] P. Papadimitratos, A. La Fortelle, K. Evenssen, R. Brignolo, and S. Cosenza. Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation. *Communications Magazine, IEEE*, 47(11):84 –95, nov 2009.

[102] Panos Papadimitratos, Giorgio Calandriello, Jean-Pierre Hubaux, and Antonio Lioy. Impact of vehicular communications security on transportation safety. In *INFOCOM Workshops 2008, IEEE*, pages 1–6. IEEE, 2008.

[103] Bryan Parno and Adrian Perrig. Challenges in securing vehicular networks. In *Workshop on hot topics in networks (HotNets-IV)*, pages 1–6, November 2005.

[104] Adrian Perrig, Ran Canetti, J Doug Tygar, and Dawn Song. The tesla broadcast authentication protocol. *RSA CryptoBytes*, 5, 2005.

[105] J. Petit, F. Schaub, M. Feiri, and F. Kargl. Pseudonym schemes in vehicular networks: A survey. *Communications Surveys Tutorials, IEEE*, 17(1):228–255, Firstquarter 2015.

[106] Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, August 2010. v0.34.

[107] Eftychios A. Pnevmatikakis, Kamiar Rahnama Rad, Jonathan Huggins, and Liam Paninski. Fast kalman filtering and forward-backward smoothing via a low-rank perturbative approach. *Journal of Computational and Graphical Statistics*, 23(2):316–339, 2014.

[108] PTV, Karlsruhe, Germany. *PTV VISSIM 5.10 User Manual*, July 2008.

[109] Fengzhong Qu, Zhihui Wu, Fei-Yue Wang, and Woong Cho. A security and privacy review of vanets. *Intelligent Transportation Systems, IEEE Transactions on*, 2015.

[110] Maxim Raya and Jean-Pierre Hubaux. The security of vehicular ad hoc networks. In *Proceedings of the 3rd ACM workshop on Security of ad hoc and sensor networks - SASN '05*, page 11, New York, New York, USA, 2005. ACM Press.

[111] Maxim Raya and Jean-Pierre Hubaux. Securing vehicular ad hoc networks. *Journal of Computer Security*, 15(1):39–68, 2007.

[112] Pieter Reyneke. A Jacobi Auction Algorithm Implementation (simple), 2012. [Online; accessed Jan-2013].

[113] S. Rezaei, Raja Sengupta, and H. Krishnan. Reducing the communication required by dsrc-based vehicle safety systems. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pages 361–366, Sept 2007.

[114] K. Sampigethaya, Mingyan Li, Leping Huang, and R. Poovendran. Amoeba: Robust location privacy scheme for vanet. *Selected Areas in Communications, IEEE Journal on*, 25(8):1569 –1589, October 2007.

[115] Krishna Sampigethaya, Leping Huang, Mingyan Li, Radha Poovendran, Kanta Matsuura, and Kaoru Sezaki. Caravan: Providing location privacy for vanet. In *in Embedded Security in Cars (ESCAR*, 2005.

[116] Florian Schaub, Zhendong Ma, and Frank Kargl. Privacy requirements in vehicular communication systems. In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 03*, CSE '09, pages 139–145, Washington, DC, USA, August 2009. IEEE Computer Society.

[117] Florian Scheuer, Karl-Peter Fuchs, and Hannes Federrath. A safety-preserving mix zone for vanets. In *Trust, Privacy and Security in Digital Business*, pages 37–48. Springer, 2011.

[118] Robert K. Schmidt, Tim Leinmüller, Elmar Schoch, Frank Kargl, and Günter Schäfer. Exploration of adaptive beaconing for efficient intervehicle safety communication. *IEEE Network*, 24(1):14–19, 2010.

[119] Elmar Schoch, Frank Kargl, Tim Leinmüller, Stefan Schlott, and Panos Papadimitratos. Impact of pseudonym changes on geographic routing in vanets. In Levente Buttyán, VirgilD. Gligor, and Dirk Westhoff, editors, *Security and Privacy in Ad-Hoc and Sensor Networks*, volume 4357 of *Lecture Notes in Computer Science*, pages 43–57. Springer Berlin Heidelberg, 2006.

[120] Raja Sengupta, Shahram Rezaei, Steven E. Shladover, Delphine Cody, Susan Dickey, and Hariharan Krishnan. Cooperative collision warning systems: Concept definition and experimental implementation. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, pages 143–155, June 2007.

[121] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In *Proceedings of the 2Nd International Conference on Privacy Enhancing Technologies*, PET'02, pages 41–53, Berlin, Heidelberg, 2003. Springer-Verlag.

[122] Steven E. Shladover and Swe-Kuang Tan. Analysis of vehicle positioning accuracy requirements for communication-based cooperative collision warning. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, pages 131–140, January 2006.

[123] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J-P Hubaux. Quantifying location privacy. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 247–262, May 2011.

[124] Reza Shokri, Julien Freudiger, Murtuza Jadliwala, and Jean-Pierre Hubaux. A distortion-based metric for location privacy. In *Proceedings of the 8th ACM Workshop on Privacy in the Electronic Society*, WPES '09, pages 21–30, New York, NY, USA, 2009. ACM.

[125] D. Streller. Road map assisted ground target tracking. In *Information Fusion, 2008 11th International Conference on*, pages 1–7, July 2008.

[126] Jinyuan Sun, Chi Zhang, Yanchao Zhang, and Yuguang Fang. An identity-based security system for user privacy in vehicular ad hoc networks. *Parallel and Distributed Systems, IEEE Transactions on*, 21(9):1227–1239, 2010.

[127] Yipin Sun, Zhenqian Feng, Qiaolin Hu, and Jinshu Su. An efficient distributed key management scheme for group-signature based anonymous authentication in vanet. *Security and Communication Networks*, 5(1):79–86, 2012.

[128] Yipin Sun, Xiangyu Su, Baokang Zhao, and Jinshu Su. Mix-zones deployment for location privacy preservation in vehicular communications. In *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, pages 2825–2830. IEEE, 2010.

[129] Yipin Sun, Bofeng Zhang, Baokang Zhao, Xiangyu Su, and Jinshu Su. Mix-zones optimal deployment for protecting location privacy in VANET. *Peer-to-Peer Networking and Applications*, jun 2014.

[130] Andreas Tomandl, Florian Scheuer, and Hannes Federrath. Simulation-based evaluation of techniques for privacy protection in vanets. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference on*, pages 165–172. IEEE, 2012.

[131] Andrea Tomatis, Hamid Menouar, and Karsten Roscher. Forwarding in vanets: Geonetworking. In *Vehicular ad hoc Networks*, pages 221–251. Springer, 2015.

[132] Y. Toor, P. Muhlethaler, and A. Laouiti. Vehicle ad hoc networks: applications and related technical issues. *Communications Surveys Tutorials, IEEE*, 10(3):74 –88, sep 2008.

[133] G. Tóth, Z. Hornák, and F. Vajda. Measuring Anonymity Revisited. In *Proceedings of the Ninth Nordic Workshop on Secure IT Systems*, pages 85—-90, 2004.

[134] M. Ulmke and W. Koch. Road-map assisted ground moving target tracking. *Aerospace and Electronic Systems, IEEE Transactions on*, 42(4):1264–1274, October 2006.

[135] Sandesh Uppoor and Marco Fiore. Vehicular mobility trace of the city of cologne, germany, 2011. [Online; accessed 20-January-2015].

[136] Sandesh Uppoor, Oscar Trullols-Cruces, Marco Fiore, and Jose M. Barcelo-Ordinas. Generation and analysis of a large-scale urban vehicular mobility dataset. *IEEE Transactions on Mobile Computing*, 13:1061–1075, 2014.

[137] A. Wasef, Rongxing Lu, Xiaodong Lin, and Xuemin Shen. Complementing public key infrastructure to secure vehicular ad hoc networks [security and privacy in emerging wireless networks]. *Wireless Communications, IEEE*, 17(5):22 –28, october 2010.

[138] Albert Wasef and Xuemin (Sherman) Shen. Rep: Location privacy for vanets using random encryption periods. *Mob. Netw. Appl.*, 15(1):172–185, February 2010.

[139] Yu-Chih Wei and Yi-Ming Chen. Safe Distance Based Location Privacy in Vehicular Networks. In *2010 IEEE 71st Vehicular Technology Conference*, pages 1–5. Ieee, 2010.

[140] G. Welch and G. Bishop. An Introduction to the Kalman Filter: SIGGRAPH 2001 Course 8. In *Computer Graphics, Annual Conference on Computer Graphics & Interactive Techniques*, pages 12–17, 2001.

[141] Alan F Westin. *Privacy and freedom*. Atheneum, first edition edition, 1970.

[142] R. Wiedemann. *Simulation des Strassenverkehrsflusses*. Schriftenreihe des Instituts fuer Verkehrswesen der Universitaet Karlsruhe; 8. Instituts fuer Verkehrswesen der Universitaet Karlsruhe, 1974.

[143] B. Wiedersheim, Zhendong Ma, F. Kargl, and P. Papadimitratos. Privacy in inter-vehicular networks: Why simple pseudonym change is not

enough. In *Wireless On-demand Network Systems and Services (WONS), 2010 Seventh International Conference on*, pages 176 –183, Feb. 2010.

[144] T.L. Willke, P. Tientrakool, and N.F. Maxemchuk. A survey of inter-vehicle communication protocols and their applications. *Communications Surveys Tutorials, IEEE*, 11(2):3–20, Second 2009.

[145] Xinzhou Wu, Sundar Subramanian, Ratul Guha, Robert G. White, Junyi Li, Kevin W. Lu, Anthony Bucceri, and Tao Zhang. Vehicular communications using DSRC: Challenges, enhancements, and evolution. *IEEE Journal on Selected Areas in Communications*, 31(9):399–408, 2013.

[146] Yong Xi, Kewei Sha, Weisong Shi, L. Schwiebert, and Tao Zhang. Enforcing privacy using symmetric random key-set in vehicular networks. In *Autonomous Decentralized Systems, 2007. ISADS '07. Eighth International Symposium on*, pages 344–351, March 2007.

[147] S. Yaakov Bar-Shalom, 2nd Peter K. Willett, and 3rd Xin Tian. *Tracking and Data Fusion: A Handbook of Algorithms*. YBS Publishing, April 2011.

[148] Chun Yang, M. Bakich, and E. Blasch. Nonlinear constrained tracking of targets on roads. In *Information Fusion, 2005 8th International Conference on*, volume 1, pages 8 pp.–, July 2005.

[149] Bidi Ying, Dimitrios Makrakis, and Hussein T Mouftah. Dynamic mix-zone for location privacy in vehicular networks. *Communications Letters, IEEE*, 17(8):1524–1527, 2013.

[150] S. Yousefi, M.S. Mousavi, and M. Fathy. Vehicular ad hoc networks (vanets): Challenges and perspectives. In *ITS Telecommunications Proceedings, 2006 6th International Conference on*, pages 761–766, June 2006.

[151] Bin Zan, Zhanbo Sun, Macro Gruteser, and Xuegang Ban. Linking anonymous location traces through driving characteristics. In *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*, CODASPY '13, pages 293–300, New York, NY, USA, 2013. ACM.

[152] Hui Zang and Jean Bolot. Anonymization of location data does not work: A large-scale measurement study. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, MobiCom '11, pages 145–156, New York, NY, USA, 2011. ACM.

[153] Sherali Zeadally, Ray Hunt, Yuh-Shyan Chen, Angela Irwin, and Aamir Hassan. Vehicular ad hoc networks (vanets): status, results, and chal-

lenges. *Telecommunication Systems*, pages 1–25, 2010. 10.1007/s11235-010-9400-5.

[154] K. Zheng, Q. Zheng, P. Chatzimisios, W. Xiang, and Y. Zhou. Heterogeneous vehicular networking: A survey on architecture, challenges and solutions. *Communications Surveys Tutorials, IEEE*, PP(99):1–1, june 2015.