SECURING LOCATION PRIVACY IN VEHICULAR APPLICATIONS AND COMMUNICATIONS

by

GEORGE P. CORSER

A dissertation submitted in partial fulfillment of the requirements for the degree of

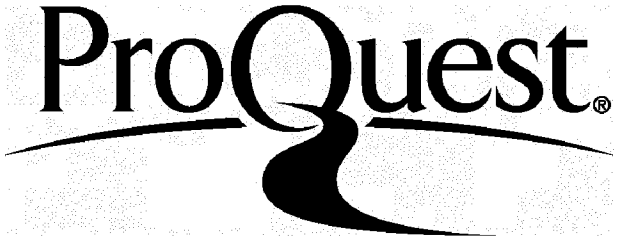
DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE AND INFORMATICS

2015

Oakland University
Rochester, Michigan

DOCTORAL ADVISORY COMMITTEE:

Huirong Fu, Ph.D., Chair
Jie Yang, Ph.D.
Jia Li, Ph.D.
Daniel Steffy, Ph.D.



To my parents,
Maureen Corser and George Albert Corser,
and to my wife, Deokhee Kim Corser

ACKNOWLEDGMENTS

The contributions of this research would not have been possible without the guidance of my academic advisor, Dr. Huirong Fu. She showed me how to take a critical view of my work, and how to incrementally improve it. I owe a debt of gratitude to Dr. Tao Shu for taking time to explain to me one-on-one the art and science of getting published. My thanks also extend to Abdelnasser Bani Hani who helped with the formatting and running some of the simulations. I wish I could thank by name the many anonymous reviewers for their constructive comments and suggestions. Finally, I innumerable thanks go to my wife, Deokhee Corser, my mother, Maureen Corser, my father, George Albert Corser, my brother, John Kevin Corser, my sister, Carin Corser, my daughter, Laura Corser, my elder son, George Ethan Corser, and my younger son, John Philip Corser. They all put up with me whenever I droned on about my dissertation, and offered challenging and constructive perspectives.

George P. Corser

ABSTRACT

SECURING LOCATION PRIVACY
IN VEHICULAR APPLICATIONS AND COMMUNICATIONS

by

George P. Corser

Advisor: Huirong Fu, Ph.D.

Vehicular communications systems may one day save lives, reduce fuel consumption, and advance connectivity, but they may also transmit information which could be deanonymized to obtain personal information. Vehicle location data are of special concern because they could be used maliciously. This dissertation presents a systematic study resulting in novel definitions, metrics and methods for evaluating and applying location privacy preserving protocols specifically in vehicular settings.

Previous work in vehicular network privacy has not thoroughly considered vehicular mobility patterns. Previous work in vehicular network privacy has not solved the problem of collusion between MAC layer and application layer attackers. As defenses against location privacy attacks, previous work has favored the privacy methods of anonymization and obfuscation, but these methods have weaknesses. Spatial-temporal cloaking, for example, requires overhead of trusted third parties, and provides little protection in low vehicle densities especially when applications require frequent precise location data. Little published work has addressed the "location" part of location privacy, the geographical distance of location privacy, focusing instead on the size of the anonymity set. The need for new metrics is indicated.

The present research addresses these issues. In addition to new definitions and metrics, this study develops privacy methods which would (1) accommodate vehicular mobility patterns, (2) defend against collusion by MAC and application layer attackers, (3) produce privacy solutions which depend on cooperation neither by large numbers of other motorists nor by trusted third parties, and (4) function in low vehicle densities, notably during the transition period between system initialization and full saturation, (5) provide protection even when applications require frequent and precise location queries, and (6) provide protection over a geographical range beyond a vehicle's wireless communications range and provide protection over measurable and lengthy spans of time. Finally, it presents a new metric for measuring privacy (KDT), an equation to estimate the safety impact of privacy protocols (SSTE), and three new privacy models, Endpoint Protection Zones (EPZ), Privacy by Decoy (PBD) and Random Rotation of Vehicular Trajectory (RRVT).

TABLE OF CONTENTS

TABLE OF CONTENTS—Continued

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AS	Anonymity Set
BI	Blind Issuer
CA	Certification Authority
CAD	Combinatorial Anonymity Degree
CPL	Continuous Precise Location
CRL	Certificate Revocation List
DSRC	Dedicated Short Range Communications
EAD	Entropy Anonymity Degree
EPZ	Endpoint Protection Zone
FCC	Federal Communications Commission
FPL	Frequent Precise Location
FPLQ	Frequent Precise Location Query
GLRP-2	Group Leader Relay Point with Members Joining in Pairs
GPA	Global Passive Adversary
IoC	Internet-of-Cars
IoT	Internet-Of-Things
IoV	Internet-of-Vehicles
LBS	Location Based Service
LD	Long-term Disclosure
LPRs	License Plate Readers
MAC Layer	Media Access Control Layer

OBU	On Board Unit
OTFP	On-The-Fly Point
PARROTS	Position Altered Requests Relayed Over Time and Space
PBD	Privacy-By-Decoy
PII	Personally Identifiable Information
PII	Personally Identifiable Information
PKI	Public Key Infrastructure
PSID	Provider Service Identifier
RRVT	Random Rotation of Vehicle Trajectories
RSU	Road Side Unit
SD	Short-term Disclosure
SMP	Stationary Mix Point
SSTE	Safety-Silence Tradeoff Equation
TAC	Traceable Anonymous Certificate
TMSs	Traffic Management Systems (TMSs)
TTP	Trusted Third Party
TTP	Trusted Third Party
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
VANET	Vehicular Ad-hoc Networks
VIN	Vehicle Identification Number

WAVE	Wireless Access in Vehicular Environments
WSMP	WAVE Short Message Protocol
ZRK	Zone-based Receiver *k*-anonymity

CHAPTER ONE

INTRODUCTION

This research addresses a general vehicular network location privacy problem: To what extent can a vehicle protect itself against surveillance even while querying a location based service (LBS), which may require frequent hyper-accurate location data? Let us call this the *frequent precise location query (FPLQ) problem.*

It may be intuitive to think of protecting the privacy of a human being, a driver, rather than the privacy of a vehicle, a car. However, it has been shown that vehicles, particularly cars, are devices as personal as mobile phones. In other words, if you can locate a person's car, you can generally locate the person, at least while the car is turned on. Futher, the techniques discussed in this research are performed by devices, not people. The FPLQ problem is after all an Internet-of-Things (IoT) problem, not a human problem, and so this work discusses protecting the vehicle, not the person, though the intent of the privacy methods is to provide benefit to humans.

The vehicular ad-hoc network (VANET) presents location privacy challenges of special complexity. In the United States VANET standards are specified by Dedicated Short Range Communications / Wireless Access in Vehicular Environments (DSRC/WAVE). These standards call for media access control layer (MAC layer) transmissions of precise vehicle locations 10 times per second. DSRC can be used to access the Internet, including LBS applications which may also require frequent precise location (FPL) data.

The privacy issue is not moot because of E911. It is possible for motorists to confuse traffic monitors even if they own mobile devices by simply turning them off or by switching devices with other people. US DOT by requiring VANETs may not allow drivers to turn systems off except for specific privacy purposes for which provisions have been made in IEEE 1609.2. Moreover, the Principle of Least Privilege remains a well-respected, fundamental and enforceable information security policy. It dictates that privacy defenders cannot fail to protect location data from one attacker simply because another potential attacker has access. Transportation monitors would work on DSRC systems completely independently from the E911 system. Transportation monitors should only be given access to such data if it is deemed important by whoever sets up the transportation system. The same is true for transportation-oriented LBSs, such as Waze or Google Navigation. If LBS owners were to allow administrators to casually monitor users, the LBS may be exposed to legal risks. Plus, there may be a moral issue, a business perception issue.

Without privacy protections in place, wireless eavesdroppers or malicious LBS administrators could track specific vehicles, or cross-reference vehicles' precise origin and termination geographical coordinates with home and work addresses, perhaps revealing (deanonymizing) a vehicle at a given location at a given time. Because motor vehicles tend to move at high speeds in predictable patterns along prescribed routes, their mobility patterns may make vehicles more vulnerable to location privacy attacks than, say, pedestrian mobile phone users. Deanonymization could occur at either the MAC layer or higher layers. MAC layer VANET systems require vehicles to transmit FPL. LBS applications sometimes have similar requirements. If there is collusion, MAC layer

data could be used to circumvent application layer (APP layer) protections such as spatial-temporal cloaking.

The vehicular location privacy problem is important because driver location data can be misused. With the assistance of hackers or malicious insiders, employers might monitor an employee's car parked at a competitor's office (revealing an employee's job interview) or at specialized medical facilities (revealing an employee's health condition). Eavesdroppers could even monitor law enforcement vehicles. It is not difficult to construct privacy breaches arising from vehicle surveillance by spouses and ex-spouses, or paparazzi and other stalkers. In fact, the problem is important enough to have been recognized by national legislators. Proposed national level legislation in the United States to address digital location privacy threats includes the Location Privacy Act and the Geolocation Privacy and Surveillance Act.

The location privacy challenge from a technical standpoint is large-scale and complicated in VANETs. Equipment supporting wireless/wifi networks is already being installed in new vehicles. Industry representatives estimate that 90% of vehicles will be wifi-connected within the decade [1]. LBS usage continues to grow rapidly [2] and is expected to expand to VANET platforms [3]. Standards governing VANETs [4] provide data fields for future privacy protocols, but the specifics of these protocols remain open research questions.

Spatial cloaking has been a standard solution to the LBS location tracking problem. The idea is, if k LBS users are operating in a spatial area, s, then k,s-privacy (a derivative form of k-anonymity [26]) is achieved [14]. The problem is, if LBS requests are repeated frequently over time, and only one of the k LBS users is consistent

throughout the set of cloaked requests, then that user is exposed. See Figure 1.1. Researchers have modified spatial cloaking to preserve *k*-anonymity even when LBSs receive frequent requests. However, no research has been performed which addresses the following problems. First, cloaking requires a trusted third party (TTP) or cloaking proxy, which may be unnecessary additional overhead. Second, cloaking is ineffective in low vehicle densities, especially if only one user is using LBS in the given vicinity.

Perhaps because of the success of cloaking, other privacy methods remain relatively under-researched. In vehicular settings for example, *dummy event* and *active decoy* methods may be effective. A dummy event is a message containing false data, sent in order to help conceal a genuine message. Dummy events and genuine messages are sent by the same entity. Dummy events function analogously to aircraft countermeasures, such as flares. An active decoy, on the other hand, is false data sent by an entity other than the entity seeking to be camoflauged. This research includes an examination of the tradeoffs between safety and privacy using dummy event and active decoy methods.

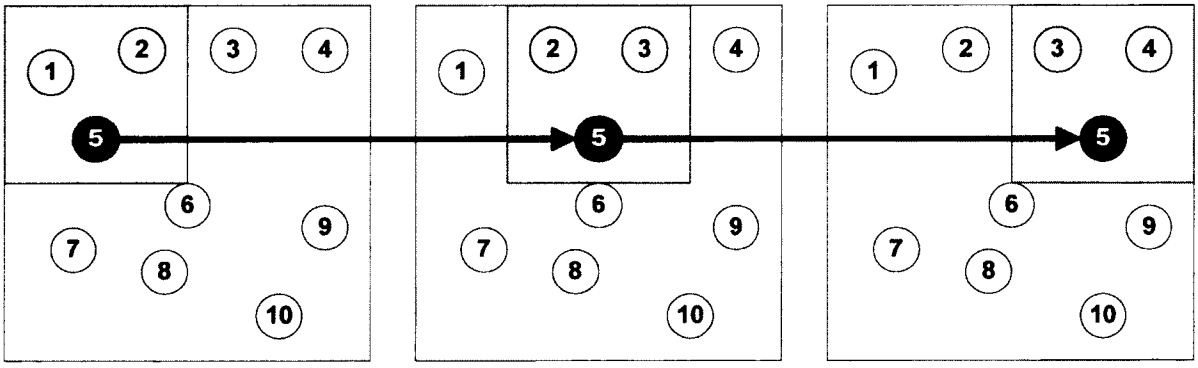


Figure 1.1: Spatial cloaking in a series of three snapshots. Vehicle 5 maintains k,s-privacy at each snapshot but all three snapshots analyzed together may reveal the vehicle.

The literature refers to several forms of privacy. *Identity privacy* is sometimes referred to as anonymity, pseudonymity or unlinkability with personally identifiable information (PII). This is often achieved by the use of pseudonyms. *Location privacy* refers to the unlinkability of PII with a geographical position, and further, the unlinkability of one pseudonym with another by using location data. *Query privacy* would make unlinkable to the user's PII, not only the location of the user, but also the particular query made or query service used. Privacy must be constrainable, as in the cases of *conditional privacy* and *revocability*. This research focuses on location privacy.

1.1 Network Model and Threat Model

Vehicles in VANETs can communicate vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I). When contacting LBS vehicles communicate V2I through a wireless access point called a roadside unit (RSU). Figure 1.2 shows how vehicles may communicate with LBS via multi-hop V2V. The present research excludes other forms of communication, such as cellular and WiMax.

Powerful RSU system administrators may have access to one, many or all RSUs. If LBS administrators have access to RSU data, LBS and RSU location data could be cross-referenced, as shown in Figure 1.2.

1.2 Vehicle Safety

Vehicle safety has historically been a matter of crash mitigation. Safety belts, air bags, collapsible steering wheels and other technologies have been designed to reduce the severity of the consequences of a crash.

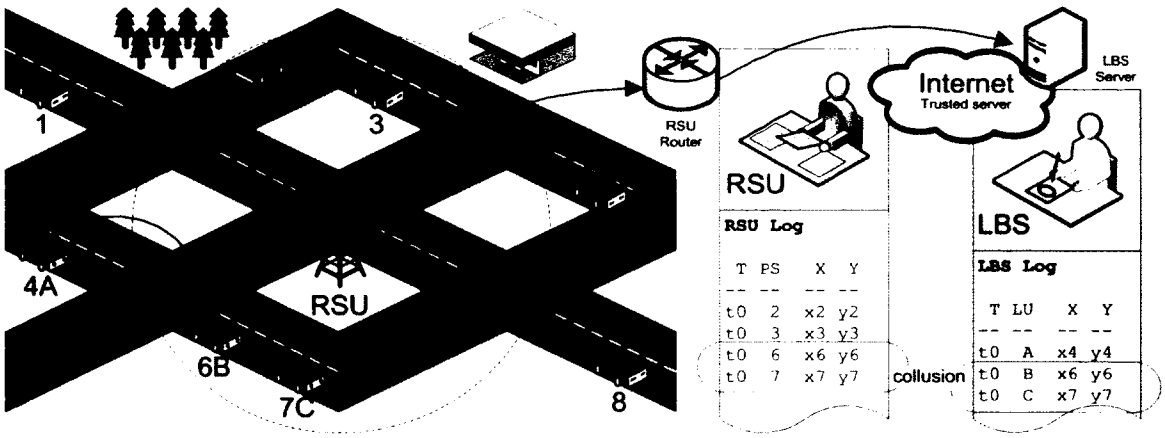


Figure 1.2: V2V and V2I Communication Scenarios.

Crash mitigation as a field of study seems to have reached a plateau. A new direction in automotive safety has arisen in the area of crash prevention. Vehicle networks have been proposed which would allow every car to compute its trajectory and the trajectories of other vehicles to alert drivers regarding potential crashes faster than human response alone would achieve. The US DOT reports such technology could eliminate 80% of crashes of unimpaired motorists [30].

Besides saving lives, crash prevention technologies such as those predicted in vehicle networks, if effective, may reduce the price of cars and insurance. Expensive crash mitigation components, like airbags, may become unnecessary, and may be superseded by more cost-effective crash avoidance components.

1.3 Location Privacy Theory and Implications

Shokri [10] suggests there are four methods for preserving location privacy: hiding events, adding dummy events, obfuscation and anonymization. See Figure 1.3. Methods can be used individually or in combination to develop defenses.

Methods vary in effectiveness in vehicular settings at the MAC and APP layers. As shown above some prior techniques have proven ineffective against APP layer queries requiring continuous precise location (CPL), a.k.a. frequent precise location (FPL).

1.4 Related Work

The literature relevant to vehicular location privacy research can generally be divided into two areas: papers which present MAC layer solutions to narrow vehicle-network-specific technical challenges, and papers which present broadly applicable theoretical concepts of location privacy. The vehicle-network-specific literature largely ignores dummy event solutions. The broader theoretical literature ignores dummy event

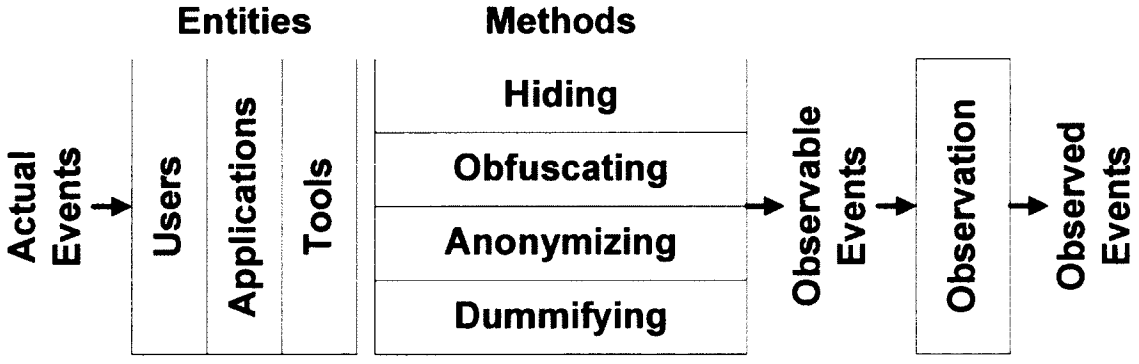


Figure 1.3: Location privacy preserving mechanisms.

solutions at least partly because of the Privacy Universality assumption, the idea that all or most users desire privacy. However, if privacy is desired by only a small subset of people, and not by all or many people, then it may be erroneous to conclude that dummy events would cause excessive overhead, even in pervasive computing environments.

It is difficult to create realistic dummies. One paper [11] reported this problem when trying to use a trajectory database of vehicle movements. No literature to date, except [50], has examined the possibility of using "live" geographical coordinates for LBS FPL queries, yet these coordinates are readily available to vehicular LBS users.

At least six previous works have studied the use of dummy events in protecting location privacy. See Table 1.1. Some [11] suggest dummy vehicle locations be generated by adding noise to traces from a trip planner. Another [12] proposes generating dummy locations in the same neighborhood as the genuine current location. Still another [13] recommends generating dummy locations entire trajectories, full trips, using algorithms which offer realistic vehicular mobility modeling and derive positions from databases of previously-recorded real driver locations. One paper [14] advocates dummy location generation using either a local grid called a virtual grid, similar to [12], or a virtual circle, which ensures k,s-privacy. The paper in [15] recommends a scheme which randomly generates dummy locations rotating with movement patterns that consistent with observed human movement, though not vehicle movement. Researchers in [16] studied dummies in packet routing, not directly relevant to this study.

None of these studies except [13] generate dummies from realistic vehicular mobility models. None consider the threat model in which LBS administrators collude with RSU administrators. None use active decoys, i.e. false locations (dummy events) of

real vehicle locations in real time transmitted by vehicles other than the target vehicle. Besides the Endpoint Protection Zone (EPZ) model [28] and Privacy-by-Decoy (PBD) model [50] we are aware of no study to date which has examined the deanonymization of endpoints in VANETs under LBS/RSU collusion.

In sum, a new body of research is needed to fill the present gap between desired vehicle network properties and existing solutions. It is possible that some location privacy methods, such as dummy events, may be more applicable than originally assumed and should be investigated further. It is the goal of the present study to begin to fill this gap.

Table 1.1

Location Privacy Techniques Using Dummy Events

Authors	**Methods**	**Category**
You, Peng and Lee (2007) [15]	Random trajectory	Spatial shift
Lu, Jensen and Yiu (2008) [14]	Virtual grid, virtual circle	Spatial shift
Chow and Golle (2009) [11]	Google Maps poly line	Trajectory database
Kido, Yanagisawa and Satoh (2009) [12]	Moving in a neighborhood	Spatial shift
Krumm (2009) [13]	Data gathered from GPS receivers, modified with noise	Trajectory database
Alnahash, et. al. (ASEE, 2014) [51]	Random trajectory confined to road grid	Spatial shift
Corser, et. al. (IEEE, 2014) [50]	"Live" dummies generated by active vehicles	Active decoy

1.5 Problem Statement

The goal of vehicular network privacy research is to provide to those who desire it the appropriate balance between (a) vehicle safety, i.e. active transmission of heartbeat messages, without silent period, (b) network efficiency, i.e. low latency, low congestion and low computational and storage overhead, (c) application availability, i.e. access to LBS even if requiring FPL, and (d) desired level of privacy or transparency.

A successful vehicular privacy system should (1) accommodate vehicular mobility patterns, (2) defend against collusion by MAC and APP layer attackers, (3) produce privacy solutions for individual motorists that do not depend on cooperation by large numbers of other motorists or by trusted third parties, and (4) function in low vehicle densities, notably during the transition period between system initialization and full saturation, (5) provide protection even when applications require FPL, and (6) provide protection over a geographical range beyond any particular vehicle's communications range.

Current solutions do not achieve the attributes listed above. Specifically, privacy literature scarcely covers dummy event solutions, which may be useful in vehicular scenarios. During the early stages of the transition period between 0% saturation and ~100% saturation of vehicle network equipment deployment, network congestion will be of lesser concern, making dummy event solutions a possible interim choice. Further, in low density situations and situations which require FPL queries by LBS, spatial-temporal cloaking will not achieve its highest privacy levels. Finally, it has yet to be determined the percentage of vehicle users who will want any privacy at all. If many prefer transparency to privacy, then dummy methods for sparse users may be a viable option.

Dummy event and active decoy methods may improve solutions to some vehicle location privacy problems. This indicates the need for additional research quantifying the tradeoffs between vehicle safety, network efficiency and privacy.

1.6 Research Objectives and Research Questions

The objectives of this work are to identify the properties of effective privacy-preserving VANETs, suggest novel protocols, evaluate the effectiveness of privacy protocols, and quantify tradeoffs between safety and privacy. Simulations address the following research questions.

- What are the properties desirable for privacy-preserving protocols?
- What solutions, with a special focus on types of dummy event solutions, might achieve some or all of these properties?
- How can these properties be measured? Especially, how can privacy be measured in terms of anonymity, distance and time? What new metrics might be devised and how would new metrics compare with previous ones?
- What effect do vehicular mobility patterns have on privacy protocols involving dummy events?
- What effect do dummy methods have on safety? That is, what duration of silent period is required for dummy methods versus prior solutions?

1.7 Methodology

This research provides theoretical analysis, formulas, and simulated results using realistic vehicle mobility patterns to evaluate privacy protocol performance in terms of vehicle safety (duration of silent period) and privacy in terms of anonymity, distance and time. To evaluate safety-privacy tradeoffs, simulations varied silent period durations and

determine how many vehicles would appear in the average anonymity set given varying vehicle densities and LBS user percentages. At the MAC layer the anonymity set would be the vehicles in the mix zone. At the APP layer it would be the vehicles in the LBS cloaking region or active decoy set.

1.8 Research Contributions

The research has yielded the following contributions.

- First, it has enumerated a consistent set of desired properties of vehicle networks and formally added undeanonymizability, efficiency and user control to the list.
- Second, it has presented new privacy metrics specific to vehicular privacy analysis, specifically *anonymity* (K), *distance deviation* (D), and *anonymity time*, (T), or KDT for short. These apply specifically to FPL and CPL.
- Third, this work resulted in a new way to use the random rotation of vehicle trajectories (RRVT) in order to protect location privacy over short distances.
- Fourth, it resulted in a new method of protecting location privacy over longer distances and lengths of time using endpoint protection zones (EPZs).
- Fifth, this work developed a way to achieve privacy by decoy (PBD), i.e. to collude with other vehicles in motion to protect privacy against very powerful adversaries.
- Sixth, the research produced an equation, the safety-silence tradeoff equation (SSTE), which serves as a way to quantify the cost/benefits of absent BSMs or silent periods on safety/location privacy of vehicles.

1.9 Outline of Dissertation

The rest of this document is organized as follows. Chapter Two enumerates desired properties of VANETs. Chapter Three describes how these properties are quantified. Chapter Four presents a new privacy model, RRVT, and evaluates its performance. Chapter Five presents a new privacy model, EPZ, and evaluates its performance. Chapter Six presents a new privacy model, PBD, and evaluates its performance. Chapter Seven presents SSTE, a mathematical equation for evaluating the cost that silent periods have on safety provided by DSRC. Chapter Eight presents KDT, a new way to measure privacy in CPL/FPL contexts. Chapter Nine concludes the dissertation and outlines a path for future work.

CHAPTER TWO

DESIRED PROPERTIES OF VANETS

First and foremost, DSRC was designed to save lives and reduce injury. Cellular communications between vehicles were deemed too slow to be used to prevent vehicle crashes. Sensors were deemed limited because they would not always function in conditions without clear line-of-sight between vehicles. Consequently, a new set of super-low-latency protocols were developed, and a new spectrum assigned specifically for vehicles. The Federal Communications Commission (FCC) dedicated a 75 MHz spectrum in the 5.9 GHz band for a new set of protocols, called DSRC in the United States [5].

DSRC features two full distinct network/transport layer protocols. See Figure 2.1. The first, WAVE, Wireless Access for Vehicular Environments, features WAVE Short Message Protocol (WSMP). WSMP would typically be used in V2V safety applications, including safety services which use SAE J2735 message protocols that include a message type called a Basic Safety Message, BSM, also referred to as a heartbeat message. The second, IPv6/TCP/UDP, would typically be used in V2I, especially when accessing infotainment applications.

DSRC was specifically designed to accommodate privacy. Unlike traditional networks, endpoints do not communicate true MAC addresses. Temporary MAC

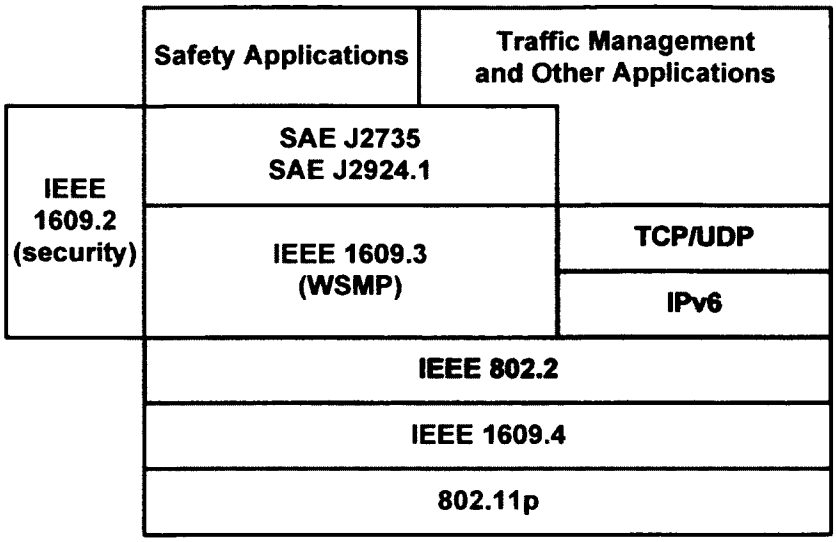


Figure 2.1: DSRC Protocol Stack. DSRC includes traditional protocols such as IPv6 and TCP/UDP, but also new protocols such as WAVE Short Message Protocol (WSMP) and SAE J2735.

identifiers, *pseudoIDs*, are switched from time to time [5]. In other words, privacy capabilities exist in DSRC that are not present in traditional networks.

As research into the implementation of VANETs has advanced, researchers have continued to identify and examine desired properties of vehicle networks. In the early days, researchers focused almost exclusively on safety. Today security features involving trust and privacy have been growing in popularity.

This chapter examines twelve desired properties, concentrating especially on their relevance to location privacy. Each property below is defined and described, including the method by which it was measured in simulations.

2.1 Property 1: Collision Avoidance

BSMs must maximize safety.

To achieve collision avoidance, vehicles inform each other regarding whether or not they are on a trajectory to collide. In VANETs this is accomplished using BSMs. BSMs must be fast, frequent, and localized, and include the transmitting vehicle's precise position, velocity and direction at precise times. For BSMs, fast means ultra low-latency wireless communication using 802.11p which omits BSS and associated MAC sublayer overhead [5].

Privacy protocols diminish safety when they include a silent period, i.e. a time span when no BSMs are transmitted. All of the protocols evaluated in this research were tested using a fixed, 30-second silent period. The total amount of silence can be computed by multiplying the number of vehicles anonymized by the time of silence (30 seconds). Since the models anonymized roughly the same number of vehicles and the silent period was constant the models performed equivalently with respect to their impact on safety.

2.2 Property 2: Authentication

BSMs must be trustworthy.

Vehicles must be able to trust the authenticity of each others' BSMs. To achieve authentication, BSMs must include valid digital signatures. Because digital signatures provide quick, low-overhead authentication, they have long been accepted as the most effective mechanism to ensure authentication and message integrity in vehicular

environments where nodes may often be in transmission range of each other for only a brief time [33].

Digital signatures do not provide confidentiality, but confidentiality is not important since BSM data do not contain secret information. In fact, BSM data are designed to be transparent for safety reasons, so confidentiality would be counterproductive. However, this lack of confidentiality, coupled with verification of credentials provided by the digital signature, also makes BSMs vulnerable to privacy attacks.

Digital signatures defined in IEEE 1609.2 [4] include Provider Service Identifier (PSID). The PSID indicates the application that will use the data in the payload, so the PSID is analogous to a port number in TCP. It is possible not all BSMs will contain identical PSIDs, in which case an eavesdropper might be able to track a vehicle by using PSIDs within the digital signature. All DSRC privacy protocols suffer from this particular deficiency, so we do not consider it when determining the rating.

Privacy protocols diminish authentication effectiveness when they introduce vulnerabilities or impair performance in processing or accessing digital signatures. Let us define the latter as public key infrastructure (PKI) efficiency. In each model except Group Leader Relay Point with members joining in pairs (GLRP-2) it was assumed that there would be multiple certificates per on-board unit (OBU) as in [63]. In GLRP-2 it was assumed that each vehicle belonging to a group would have that same number of certificates as the other models, plus additional group signature. The V2V communications necessary to establish the group, plus the communications necessary to

relay and convert messages to the LBS, represent additional impairment of PKI efficiency and therefore the GLRP-2 model was rated lower than the other models.

2.3 Property 3: Pseudonymity

BSMs must not reveal real identities of vehicles or owners.

To achieve pseudonymity, a type of identity privacy, BSMs must use pseudonyms, or *pseudoIDs*, each of which having a corresponding digital certificate. Except in circumstances requiring accountability or revocability, described below, pseudoIDs and their certificates are unlinkable to the vehicle identification number (VIN) of the vehicle and to the personally identifiable information (PII) of the vehicle owner.

In the literature the term, unlinkability, may refer to the inability to correlate vehicle identities with pseudoIDs, but it may also refer to the inability to correlate between multiple pseudoIDs of a particular vehicle. To avoid ambiguity, we refer to the former as pseudonymity and to the latter as either untrackability or untraceability, defined below.

Privacy protocols diminish pseudonymity when they risk linkage between pseudoID and VIN or PII. There is a natural tradeoff between authentication by digital signature and pseudonymity by pseudoID. For authentication to be fast and efficient, the fewer the pseudoIDs and certificates the better; for pseudonymity, the more the better. Consequently, for the same reason that the GLRP-2 model was rated lower than the other models in terms of Property 2, it was rated higher than the other models in terms of Property 3. That is, additional pseudonyms slow down the authentication process, but they at the same time add an additional layer of pseudonymity.

2.4 Property 4: Untrackability

PseudoIDs in currently transmitted BSMs must not be linkable to pseudoIDs in immediately preceding BSMs from the same vehicle.

If a vehicle were identified (marked), and its pseudoID linked to PII even a single time, then the vehicle could be monitored as long as its BSM used that same pseudoID. Hence this property may be appropriately referred to as real-time location privacy.

To achieve untrackability, a type of location privacy, BSMs must use multiple pseudoIDs, rotating between them frequently, on average every 5-10 minutes. A single vehicle may contain several, or even thousands of pseudoIDs, each with its own digital certificate. By periodically changing between many pseudoIDs theoretically a vehicle could only be tracked while a particular pseudoID was in use subsequent to the vehicle being marked [5].

Privacy protocols diminish untrackability when they risk linkage between current pseudoIDs and their immediately preceding pseudoIDs. There is a natural tradeoff between authentication by digital signature and untrackability. For authentication to be fast and efficient, the fewer the pseudoIDs and certificates the better; for untrackability, the more the better. If a pseudoID switching technique in a privacy protocol includes a silent period when no BSMs are transmitted, there could also be a tradeoff between collision avoidance and untrackability.

The group model provides no more barriers to tracking than the other models, because, while there more pseudonyms are involved, the ability of an eavesdropper to determine which vehicle belongs to a given signal (blip) does not depend on pseudonym. Under the assumptions in this paper, a vehicle is marked upon entering the region. No

matter what identifier the vehicle is transmitting, the geographical coordinates alone are sufficient to track it because it is transmitting frequently, every 100 ms. Any vehicle which changes identifier is only protected from tracking if there are other members in its anonymity set. Untrackability is accomplished, not by pseudo-identity, but by mixing. (Refer to the discussion on unlinkability in the prior subsection.) In sum, all models are equivalently untrackable.

2.5 Property 5: Untraceability

PseudoIDs in current or past BSMs must not be linkable to other pseudoIDs from the same vehicle, except by proper authorities.

To achieve untraceability, BSMs must be transmitted using multiple pseudoIDs, switching between them, as in untrackability, above. However, the property of untraceability is distinct from untrackability. By this paper's definition, "tracking" a vehicle would be performed in real-time, while the vehicle is in motion. "Tracing" the vehicle would be a matter of historical investigation, to determine what vehicle was at what location at what time. Hence this property may be appropriately referred to as historical location privacy.

Historical evidence-gathering has been used by proper authorities, such as courts of law (see accountability, below). But tracing could also be used by stalkers or paparazzi for gathering background information on vehicles to establish locations at specific times or to establish transportation patterns of people under unauthorized surveillance.

Privacy protocols diminish untraceability when they risk linkage between pseudoIDs and preceding pseudoIDs for a given vehicle. There is a natural tradeoff between authentication by digital signature and untraceability. For authentication to be

fast and efficient, the fewer the pseudoIDs and certificates the better; for untraceability, the more the better. If a pseudoID switching technique in a privacy protocol includes a silent period when no BSMs are transmitted, as in this study, there could also be a tradeoff between collision avoidance and untraceability.

To the extent that group signatures add an additional layer of identity privacy via pseudonymity, group signatures provide better protection against tracing.

2.6 Property 6: Accountability

PseudoIDs must be linkable to PII by proper authorities.

Sometimes it is beneficial to link a vehicle to its owner's identity and/or its location, such as when a vehicle may have been used in a crime or involved in an accident. It may be argued that a privacy protocol without the property of accountability would introduce more risk to the public by concealing criminals than it would introduce security to the public by protecting people's privacy.

To achieve accountability, a certificate authority (CA) or other trusted third party (TTP) must protect vehicle and owner identity and location while maintaining the capability to link this information with pseudoIDs if requested by a proper authority. This is sometimes referred to as conditional privacy.

Privacy protocols diminish accountability when they do not provide a secure mechanism for linkage between pseudoIDs and vehicle/owner identity and location. There is a natural tradeoff between trust and privacy, and this balance must be struck in the property of accountability. The TTP must be able to determine the circumstances under which a proper authority may circumvent a privacy protocol and reveal the true identity associated with a pseudoID.

All of the simulated protocols require identities and pseudo-identities with digital certificates issued by a CA which could deanonymize the transmission if, for example, required to do so by a court order. The group model provides for both a CA to enable single-entity digital signatures and a group manager to enable group signatures. All models provide the same technical level of accountability. It is assumed that it would be more difficult for a law enforcement authority to obtain a warrant and receive deanonymized vehicle information from two entities than from just one, so the group model is rated inferior to the others because it includes both the CA and the group leader to cooperate in order to achieve accountability.

2.7 Property 7: Revocability

PseudoIDs and digital certificates must be rescindable.

It is possible that valid digital certificates could be stolen and used maliciously. If this is detected the certificate should be revoked.

To achieve revocability, a CA or other TTP must provide valid digital certificates for pseudoIDs while maintaining the capability of rescinding certificates by updating and distributing a certificate revocation list (CRL) if requested by a proper authority.

Privacy protocols diminish revocability when they impair the distribution of CRLs securely, quickly and broadly [39]. For authentication to be fast and efficient, the smaller the CRLs, the better; for effective revocability, some protocols indicate large CRLs. There is a natural tradeoff between trust and privacy, and this balance must be struck in the property of revocability. The TTP must be able to determine the circumstances under which more harm than good comes from BSMs bearing a particular

pseudoID and that the benefit of revoking that pseudoID's digital certificate exceeds its cost.

Some researchers suggest the group model is infeasible due to limitations of bandwidth and computation power, since pseudoID schemes create large certificate revocation lists, CRLs, and associated checking costs, network overhead necessary to verify that certificates have not been revoked [9].

2.8 Property 8: Anonymity

Location privacy models must maximize indistinguishability between pseudoIDs.

Privacy protocols can be evaluated by anonymity, which we define as the quantifiable amount of privacy the vehicle's pseudoID enjoys by using the protocol. Anonymity could measure identity privacy or location privacy. The pseudoID is the mechanism which protects identity privacy, therefore it follows that pseudonym anonymity could measure identity privacy protection. But what about location privacy?

There are four methods of preserving location privacy: obfuscation, hiding events, adding dummy events, and anonymization [10]. For BSMs, obfuscation is not possible because precision is required for safety applications. Hiding BSMs is not possible because safety applications depend on the detectability of BSMs. Adding dummy BSMs may threaten safety by inducing vehicles to react to nonexistent vehicles; in fact digital signatures are used to reduce the possibility of malicious fake BSMs. The only remaining method is anonymization. Since the identities used in BSM transmissions are pseudonyms, pseudoIDs, the only way to protect privacy in VANETs is by pseudonym anonymity, or as we call it, anonymity.

It is necessary to make fine distinctions between the terms, anonymous and pseudonymous, to clarify computational privacy in vehicular network contexts. The dictionary definition of anonymous is, "not named or identified" [40], a definition which might not apply in networks that require identifiers for communications. When referring to computers the term, anonymous, may imply a pseudonym which is unlinkable to a person's true identity, as in an anonymous post on a blog. This definition introduces ambiguity with the term anonymity when used as in anonymity set, defined below. We use the dictionary definition of pseudonymous, "bearing or using a fictitious name" [41] to indicate unlinkability to PII. Vehicle networks use pseudoIDs, which achieve the property of pseudonymity. This paper uses the term anonymity as it is used in set theory, as in an anonymity set [42]. Thus we can define two distinct privacy properties, pseudonymity and pseudonym anonymity.

To maximize anonymity, location privacy models must maximize the anonymity set size of each pseudoID. Anonymity has been measured using a range of metrics, not all of which apply to the evaluation relevant to this paper. Let us set aside *l*-diversity [17], *t*-closeness [18], *L1*-similarity [44], *m*-invariance [45][46], and ε-differential privacy [19]. We also set aside network metrics such as combinatorial anonymity degree (CAD), zone-based receiver *k*-anonymity (ZRK) and evidence theory anonymity [47]. These are all valuable measures of anonymity but they do not measure pseudonym anonymity in the context of BSMs.

We evaluated pseudoID anonymity of privacy protocols using one or more of the following metrics: pseudonym anonymity set size, |AS|; entropy of the anonymity set size, H(|AS|), also called entropy anonymity degree (EAD) [47]; and tracking probability,

Pt. We evaluate each privacy protocol on its ability to achieve indistinguishability amongst pseudoIDs. Note that the term, pseudoID, can be used for the temporary MAC address at the MAC layer, or the identifier used at the APP layer, including a dummy or decoy message.

Just as digital signatures may introduce privacy vulnerabilities, as discussed in authentication, above, BSMs may introduce privacy vulnerabilities which could be used to reduce anonymity set size by isolating characteristics peculiar to a specific target vehicle. Some of the data fields in a BSM, for example, could be used by eavesdroppers to link past pseudoIDs to current ones.

If `MsgCount`, a sequence number field used for packet error monitoring, is not reset when pseudoID changes then an eavesdropper could match the sequence numbers of past BSMs with current BSMs and link the pseudoIDs. If `PositionalAccuracy` differed from vehicle to vehicle, then an eavesdropper could link pseudoIDs or even track this data element instead of tracking pseudoIDs. If `SteeringWheelAngle` were misaligned for a vehicle, or if `BrakeSystemStatus` were in any way distinct for one vehicle compared to its neighbors it make an easy mark. Perhaps the most concerning data element is `VehicleSize`, which provides vehicle length and width accurate to a resolution of 1 cm. Vehicles are produced in sufficiently various lengths and widths that this element alone may be sufficient for an eavesdropper to link past and present pseudoIDs.

All DSRC privacy protocols suffer from this particular deficiency, so we do not consider it when determining the rating, but the challenge remains: how to implement fine-grained safety data without exposing driver location.

Privacy protocols diminish anonymity when they do not provide high levels of indistinguishability (unlinkability) between pseudoIDs, as measured by (1), (2) and (3). Equations (7) and (12) measure distance deviation and anonymity time, respectively. Equations (1), (7) and (12) are used to compute the composite metric, KDT. The formulas for all of these equations appear in the following chapter.

2.9 Property 9: Decentralization

BSMs must not be traceable by a single TTP.

Decentralized protocols by definition involve multiple independent TTPs. The purpose of decentralization is to prevent a single TTP from being able to undeanonymize vehicles.

To achieve decentralization, BSMs could, for example, include blind signatures, which require a traceable anonymous certificate (TAC), a certificate issued by two TTPs, a Blind Issuer (BI) and an Anonymity Issuer (AI). TACs work the same as other digital certificates, but it would require the cooperation of multiple TTPs to trace back to the identity of the vehicle. For a full discussion, see [48] and [49].

Privacy protocols diminish decentralization when they fail to call for multiple independent TTPs. Any protocol could implement blind signatures but the group model explicitly requires at least two separate credential-issuing entities (TTPs) so GLRP-2 is evaluated higher than its peers in this regard.

2.10 Property 10: Undeanonymizability

Location privacy models must minimize cross-referenceability.

If vehicles require LBS query results based on precise location data, then eavesdroppers with access to the content of that data could use map databases or other

cross-references to deanonymize motorists and vehicles. For example, if a driver started her vehicle at coordinates *(x,y)*, her home address, then eavesdroppers could match *(x,y)* with the longitude and latitude of the address and possibly identify the driver.

Another method of deanonymization is used in this paper. It occurs when vehicles enter an ingress point of a region, or when they exit an egress point. Deanonymization may be possible by using license plate readers (LPRs) at these ingress/egress points and identifying the vehicle using vehicle registration records.

Cross-referencing could occur in real time in the case of RSU-LBS collusion. BSMs provide RSU with vehicle FPL and MAC-layer pseudo-identifier of vehicle. So once deanonymized any eavesdropper with access to RSU or OBU data in communications range of the target could continuously track the target vehicle. FPL LBS queries provide LBS with vehicle FPL and APP-layer identifier. If RSU and LBS collude then LBS could also track the target in real time. If a target vehicle were to go radio silent with respect to BSMs, but not LBS queries, LBS could continue to track the target. If a target vehicle were to go radio silent with respect to FPL LBS queries, but not BSMs, then LBS could still continue to track target using BSMs. This demonstrates the necessity that vehicles must go radio silent at both MAC-layer and APP-layer to defend against RSU-LBS collusion.

Combination MAC-layer/APP-layer defenses, such as endpoint protection zones (EPZ) [28] could enable undeanonymizability even when LBS and RSU collude. Such defenses could be used in combination with any protocol in this study, however EPZ defends endpoints only; it does not not defend against LPRs at region border ingress/egress points.

Dummy event methods [11][12][13][14][15] have been researched but none, except the On-The-Fly Point (OTFP) model, Privacy-by-Decoy (PBD) [50], has explicitly addressed the RSU-LBS collusion problem beyond the basic defense offered by EPZ. PBD not only protects against RSU-LBS collusion, but also enables LBS login. This is the strongest undeanonymizability defense of which we are aware, especially considering it requires no anonymization server. The other models offer no protection against deanonymization from cross-referencing or collusion.

2.11 Property 11: Efficiency

Location privacy protocols must minimize network and application overhead and congestion.

In order to maximize efficiency location privacy protocols must avoid overhead, unnecessary consumption of bandwidth, memory and computational resources. Network overhead can be measured by the number (or percentage) of additional messages required to implement the protocol.

2.12 Property 12: User Control

Location privacy models must maximize individual customization settings.

Not all motorists prefer privacy. Some prefer transparency. Some may prefer to switch between transparency and privacy. The capability of a location privacy model to permit end user customization or even the capability to shut off anonymization altogether, indicates a superior privacy protocol compared to one that does not offer the feature.

This is an important and often overlooked privacy property. Consider the following scenarios, all of which involve customizable location privacy.

- A political candidate may want location transparency when on the campaign trail, but may want location privacy when meeting with potential donors.
- A parolee (or a parole enforcement administration) may want provable transparency (nonrepudiation) to prove a parolee was driving to work and not at a criminal meetup.
- The owner of a fleet of trucks may not want his drivers to have location privacy configurable at the driver level, however she may not want her competitors to know her fleet's location.

CHAPTER THREE

PERFORMANCE METRICS

Now that properties have been defined we need ways to measure how well protocols perform with respect to these properties. Since this report confines itself to the FPLQ problem, it omits considerations such as the complexities of encryption or key distribution, for example. The challenge in front of us is how to keep vehicles anonymized even when they are connected to an LBS and transmitting precise location frequently.

3.1 Safety-related Properties

As shown in Figure 3.1, there are four properties involving safety. Property 1, collision avoidance, is best served when 100% of vehicles transmit BSMs 100% of the time. However, Properties 4, 5 and 10 require silent periods for switching identifiers.

It would be pointless to switch identifiers without a silent period under FPLQ because transmissions are so frequent that an attacker could easily see which prior vehicle switched to which subsequent identifier. Silent periods must be instituted otherwise ID switching could not be used to achieve anonymity.

Silent periods diminish the effectiveness of safety, but FPLQ solutions studied here require generally equal durations of silent periods. Consequently this research does not compare protocols in terms of their safety risks. Rather, it presents the overall safety risk of using any silent period method to achieve privacy. See Chapter Seven.

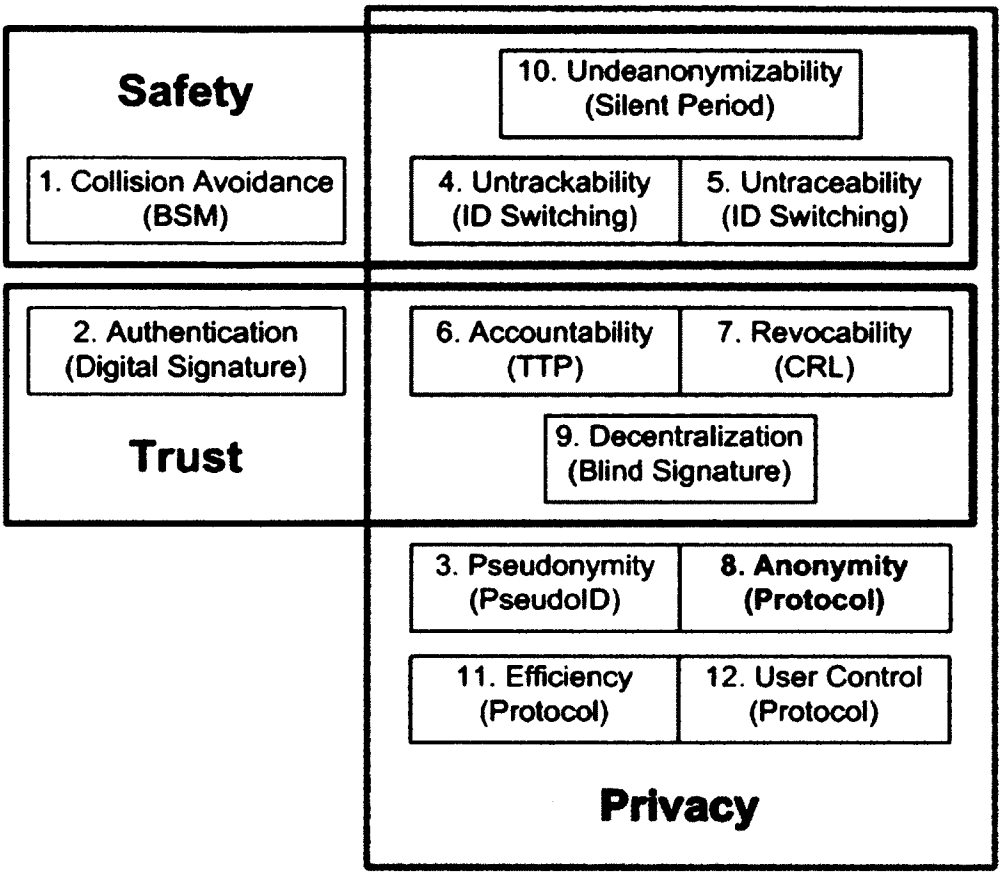


Figure 3.1: Desired properties of VANETs.

3.2 Trust-related Properties

Four of the twelve properties, 2, 6, 7 and 9, are categorized as trust-related. All of them involve how encryption keys are used, issued, or maintained. Since a vehicle's location is in the open necessarily, for safety reasons, and location is the only data with which this report is concerned, we omit metrics for measuring the relative value of the various methods, such as trusted third parties, certificate revocation lists, blind signatures, etc. All of these matters are important, but none bear directly on the FPLQ problem.

3.3 Privacy-related Properties

The remaining four properties, 3, 8, 11 and 12, concern themselves with privacy, not safety or trust, however not all of them bear directly on our present problem, FPLQ's. Property 3, pseudoID creation and usage, for example, depends on choices made by programmers and standards set by IEEE 1609.2 and SAE J2735. No particular pseudoID technique would be better or worse under FPLQ conditions.

Property 11, efficiency, meaning the efficiency of the privacy protocol, not the over all network efficiency, does have a measurable impact and we can compare effects under differing privacy protocols. See Chapter Nine for this information.

Property 12, user control, might best be evaluated using a rating scale, for example a zero-to-ten scale. Perhaps zero could indicate the user has no control; that is, the privacy protocol forces the driver's car to execute the protocol. Perhaps ten would indicate that the user could adjust on the fly when and where and how his car is anonymized. In this study, we used an on-off evaluation. Either the user had some control (on) or she had none (off).

Property 8, anonymity, may be the most measurable of the pure-privacy properties. Important work has been done in this area, and has been applied in a range of situations. However, there are some peculiarities to VANETs that suggest that even more work needs to be done. Below, see previous work and the new measurement techniques developed during this research. This dissertation expands the scope of anonymity to include distance and time, and introduces theoretical estimates and metrics to evaluate performance of this expended definition of the property of anonymity.

3.3.1 Traditional Macroscopic Location Privacy Definitions and Metrics

Microscopic location privacy is defined in [10] to be anonymity at a specific place and time. On the other hand, *macroscopic* location privacy is defined as anonymity from beginning to end of an entire trajectory, i.e. a path or segment of a path traversed by a user, in this paper, a vehicle. This research examines macroscopic location privacy.

Early VANET privacy research often focused on MAC layer location privacy problems using microscopic location privacy metrics. More recently researchers have begun to study application (APP) layer location privacy problems using macroscopic location privacy metrics. This study belongs to the latter category.

Users are vehicles, not people, because Traffic Management Systems (TMSs) may require queries with such rapidity and precision that queries are automated. TMSs therefore are herein considered Internet-of-Things (IoT), a.k.a. Internet-of-Cars (IoC) or Internet-of-Vehicles (IoV), applications.

The traditional methods can be adapted to macroscopic purposes by replacing individual entities with trajectories. That is, instead of examining an entity's anonymity at a specific point in space and time, examine its trajectory over a region of space and duration of time. This paper adapts the definitions from [58] to define the metrics below, however before embarking on that certain more basic definitions must be established, specifically definitions of privacy which apply to vehicle networks.

Some have defined privacy as user control over information [26], or the degree to which individuals can determine for themselves when, how, and to what extent location information about them is communicated [27]. IEEE 1609.2 measures privacy using anonymity [4]. (The term, anonymity, is defined as in set theory.) This paper defines

privacy, location privacy, network privacy, and continuous privacy as measurable attributes of entities, and measures privacy using anonymity, distance deviation and time.

Definition 1. ***Privacy***: the degree to which an entity cannot be linked to its identity.

Definition 2. ***Location privacy***: the degree to which a spatial characteristic of an entity cannot be linked to its identity.

Definition 3. ***Network privacy***: the degree to which an entity cannot be linked to its identity while it is connected to a communications system.

Definition 4. ***Continuous privacy***: the degree to which, over a contiguous series of time intervals, an entity cannot be linked to its identity.

Definition 5. ***Continuous network location privacy***: the degree to which, over a contiguous series of time intervals, a spatial characteristic of an entity cannot be linked to its identity while it is connected to a communications system. Continuous network location privacy protects against deanonymization using frequent precise location queries.

3.3.1.1 Anonymity Set Size. The anonymity set, AS_i, of target LBS user, i, is the collection of all LBS users, j, including i, within the set of all LBS userIDs, ID, whose trajectories, T_j, are indistinguishable from T_i, within some nonzero probability, p.

$$AS_i = \{ j \mid j \in ID,\ p(i,j) \neq 0 \} \tag{3.1}$$

3.3.1.2 Entropy of the Anonymity Set Size. Entropy represents the level of uncertainty in the correlations between trajectory T_i and trajectories T_j. For background on this metric, see [25]. The entropy H_i of AS_i is:

$$H_i = -\sum_{j \in AS_i} p(i,j) \times log_2(p(i,j)) \tag{3.2}$$

3.3.1.3 Tracking Probability. Tracking probability, Pt_i, is the probability that the anonymity set size of a vehicle, as defined in (1), is equal to one, which can be written as follows.

$$Pt_i = Pr(|AS_i| = 1) \tag{3.3}$$

If $AS_i = k = 1.0$, then a vehicle has no anonymity. To measure a system's overall tracking probability, compute the percentage of vehicles with $Pt = 1$. For example, if 47% of all vehicles have $Pt = 1.0$, then one might report that a system assures 53% anonymity.

One problem with all three above metrics is they do not account for changes in the anonymity set size (AS size) over the course of a vehicle's trajectory. It is possible, even probable, that one or more vehicles in an anonymity set (AS) may become deanonymized, leaving the remaining vehicles in that AS with a reduced AS size during their trajectories. Using only the beginning AS size for all trajectories might overestimate the overall anonymity level of a system. Using only the ending anonymity set size might underestimate it.

Moreover, the above equations do not quantify spatial dispersion, the span of distance over which vehicles are spread, which prima facie seems critical for evaluating location privacy. For example, an AS spanning 10 meters may not provide the same location privacy as one spanning 10 miles. Researchers in [19] presented a metric which considers distance deviation over a trajectory, but it measures the distance between locations, not between vehicles, so their equation does not perfectly apply to the conditions considered in this paper.

Finally, the above performance metrics do not consider the duration of anonymity. Privacy protocols perhaps should be considered more effective the longer they last.

3.3.2 Proposed Metrics

This section presents a composite metric, *KDT*, for evaluating location privacy provided by privacy-preserving protocols in contexts which involve FPL queries. The composite metric uses anonymity quantification based on *k*, called k_j, the anonymity set size at a specific point in time, t_j. It uses a Euclidean distance quantification, $\bar{d}_j$, the degree of spatial dispersion, or average span of distance, between members of an anonymity set at a specific point in time, t_j. And it uses a new metric, *T*, anonymity duration, which is the number of contiguous time intervals after t_0 which have anonymity set sizes greater than one. (The first time interval with anonymity set size greater than one is labeled t_0.) The statics metrics, *k, d, and t*, quantify privacy at specific points in time. The composite metric, *KDT*, quantifies the average privacy of a system over a span of time.

3.3.2.1 Anonymity Duration (| T |). Anonymity duration, | T |, is the size of the set, T, of time intervals, t_j, j=0,1,2,..., | T |, over which an entity enjoys continuous network location privacy. So T is a collection of contiguous time intervals starting from t_0 during which the anonymity set size, $| AS_j |$, is greater than one. Formally,

$$T = \{ t_j | (t_j = t_0) \vee [(t_{j-1} \in T) \wedge (| AS_j | > 1)] \} \tag{3.4}$$

3.3.2.2 Average Anonymity Set Size (K). Anonymity set size, k_j, is the number of entities that might be confused with one another at time, t_j, that is, $k_j = |\ ASj\ |$. Average anonymity set size, K, is the sum of all k_j from t_0 to t_j divided by $|\ T\ | + 1$. Formally,

$$K = \frac{k_0 + k_1 + k_2 + \cdots + k_{|T|}}{|T| + 1} \tag{3.5}$$

3.3.2.3 Average Distance Deviation (D). The distance, d_{sij}, between two entities, s and i, in time interval t_j, is, $d_{sij} = \sqrt{(x_{sj} - x_{ij})^2 + (y_{sj} - y_{ij})^2}$. Let p_{sij} be the probability that an attacker will guess that entity i is the target, given that entity s is the actual target, in time interval t_j. The average distance deviation at time t_j is the weighted sum,

$$\bar{d}_j = \frac{1}{k_j} \frac{1}{k_j} \sum_{s=1}^{k_j} \sum_{i=1}^{k_j} p_{sij}\, d_{sij} \tag{3.6}$$

The average distance deviation over all time intervals is,

$$D = \frac{\bar{d}_0 + \bar{d}_1 + \bar{d}_2 + \cdots + \bar{d}_{|T|}}{|T| + 1} \tag{3.7}$$

3.3.2.4 Expected Average Anonymity Set Size (E[K]). If a vehicle belongs to an anonymity set as described in (1), and there are $k - 1$ other vehicles in the set, then by definition $|\ AS\ | = k$ for all vehicles in AS. To estimate k, assume that vehicles arrive at an anonymization point, P, during a time interval, w, according to a Poisson process, a common practice. See [16] and [17] for examples. Let random variable $W = w$ be the fixed time interval during which vehicles arrive within range of P. Let the inter-arrival time between vehicles have an exponential distribution with a mean of $1/\lambda$. Let X be a random variable, the number of vehicles that arrive within range of P during time W. Then the probability that $X = x$ at $W = w$ can be written as shown in (3.8). The expected value of X is shown in (3.9). The expected value of X is the expected value of k.

$$\Pr[X = x \mid W = w] = \frac{(\lambda w)^x}{x!} e^{-\lambda w} \tag{3.8}$$

$$E[X|W = w] = \sum_{x=0}^{\infty} x \Pr[X = x \mid W = w] = \lambda w \tag{3.9}$$

So, if vehicles arrive at P at a rate of one every five seconds, i.e. $\lambda=\frac{1}{5}$, in a given 30-second time interval, i.e. w=30, then the expected value of k = E[k] = E[X] = (1/5)(30) = 6. By linearity, the expected value of K is the average of all values of k, which is k, i.e. E[K] = k.

3.3.2.5 Expected Average Distance Deviation (E[D]). It is possible to estimate distance deviation, d, between two vehicles when certain attributes of the roadway topology are assumed.

3.3.2.5.1 One Point in Time, One Vehicle in Motion (E[d]). One way to estimate theoretical distance deviation is to identify an anonymization point and compute the expected distance between the anonymization point and some deanonymization point of interest. A unit square region with straight roadways provides mathematically clear illustration and generally correlates with topologies of government jurisdictions and roadways.

Assume two vehicles anonymize at the centroid, C, of a square region, R, but that only one of them continues in motion, and that the motion is at a constant rate of speed in a straight line. Assume the moving vehicle has an equal probability of driving from C to any boundary point, B, on the square. What is the expected distance traveled by the moving vehicle?

All four sides of the square are line segments of equal size, so C and the endpoints of the line segments form, not only similar triangles, but identical triangles, except for

rotation. That is, the average distance from the centroid to one line segment is identical to the average distance from the centroid to any other line segment on the square.

Let the square be a unit square centered at (0.5,0.5) with the x-axis boundary spanning from (0,0) to (1,0). The probability density function of a vehicle traveling from the centroid to any point on the x-axis boundary conforms to the uniform distribution.

$$f(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & otherwise \end{cases} \tag{3.10}$$

The expected distance a vehicle would travel after having anonymized at the centroid, C, would be the sum of the products of the probability of traveling from C to a certain boundary point, (x, y), and the Euclidean distance between the centroid and that boundary point. Since y always equals zero on the x-axis boundary the distance from C to (x, y) is

$$dist(C,B) = \sqrt{(x-0.5)^2 + (0-0.5)^2} \tag{3.11}$$

To compute the expected distance deviation, sum all possible products by integrating as follows.

$$\begin{aligned} E[dst] &= \int_{-\infty}^{\infty} f(x)\, dist\,(C,B)\; dx \\ &= \int_0^1 \sqrt{(x-0.5)^2 + 0.25}\; dx \end{aligned} \tag{3.12}$$

Computing the definite integral above, the expected distance, then, is roughly 0.5739 [57], which is the expected distance deviation from a stationary vehicle remaining at the centroid to a moving vehicle at the time it crosses the boundary of a unit square region.

3.3.2.5.2 Full Trajectory, Two Vehicles in Motion (E[d]). What if attacker and defender wish to know the expected distance deviation, not just at the square's boundary, but over a vehicle's trajectory? With some assumptions it is possible to formulate an estimate.

Assume the square region, R, has straight roadways intersecting at perpendicular angles, but that the roadways are not perpendicular or parallel to the sides of the square. Assume the same anonymization point, the centroid, C, which is also the intersection of two straight roads. Assume the distance from the centroid to the endpoint of the straight line road segment is *E[dst]* as computed in (3.10). See Figure 3.2.

Assume the two members of the anonymity set begin traveling at the same point in time starting at the same intersection. Finally, assume that both vehicles travel at the same rate of speed, r, and there is an equal probability that they will proceed in each of the intersection's four possible directions.

Then there are four possibilities. First, if the decoy vehicle travels along the same road as the target vehicle, then d=0 at all points during the trajectory. Second, if the decoy vehicle travels in the opposite direction of the target vehicle, then d=2l, where l is the length of road traveled by one vehicle at any given point in time. Third (and fourth), if the decoy vehicle travels along one of the perpendicular roads, then d=$l\sqrt{2}$, where l is the length of road traveled by one vehicle at any given point in time. There is a 25% probability of each of the four scenarios. The expected value of d, then, would be as follows.

$$\begin{aligned} E[d] &= (0.25)\int_0^{dst}(0 + 2l + l\sqrt{2} + l\sqrt{2})dl \\ &= (0.25)\int_0^{dst}(2l + 2l\sqrt{2})dl \end{aligned} \tag{3.13}$$

E[d] evaluates to approximately 0.1988 [57].

3.3.2.6 Expected Anonymity Duration (E[T]). By linearity, the expected anonymity time is the expected distance of a full trajectory, divided by the constant movement rate, *r*. Using terms from (3.12) we can write as follows.

$$E[|T|] = \frac{E[D]}{r} \tag{3.14}$$

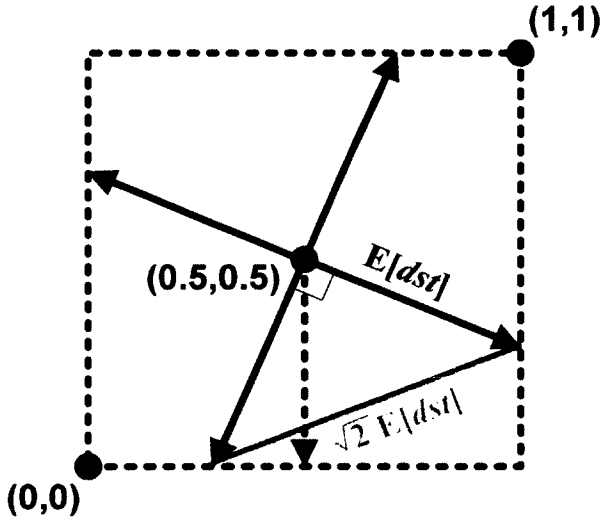


Figure 3.2: Expected distance diagram. E[dst] is the expected distance one vehicle would travel from the centroid to the boundary of a square region. The dotted square represents the region, R, a unit square. The solid perpendicular lines, each starting at (0.5,0.5) and terminating at the boundary of the square, represent road segments.

3.4 Summary

Vehicles travel fast, often on very predictable paths. In order for privacy to be effective, we hypothesize, they not only must be anonymized amongst other vehicles but there must be significant distance between targets and other members of the anonymity set. After all, if all the vehicles in an anonymity set are platooned like ducks in a row, then the distance range of the target's anonymity is limited. And if a target must frequently go silent to change identifiers this limits the benefits of both safety from BSMs and accessibility to the TMS. The results from new metrics are presented in Chapter Eight.

CHAPTER FOUR

RANDOM ROTATION OF VEHICLE TRAJECTORIES (RRVT)

Perhaps the simplest dummy event technique in location privacy models is the *direct spatial shift*. An example of this technique would be to send an LBS genuine coordinates, (x,y), and dummy coordinates, $(x+\Delta x, y+\Delta y)$. After the LBS returned its two results the sender would process those corresponding to (x,y), and discard those corresponding to $(x+\Delta x, y+\Delta y)$. Meanwhile the LBS could not be 100% certain which coordinates were genuine.

Direct spatial shift fails when the LBS knows the road map, i.e. which coordinates correspond to roadway positions. If $(x+\Delta x, y+\Delta y)$ does not fall on a roadway, then the LBS would know the genuine position of the sender. This form of deanonymizing a vehicle's position is called *map deanonymization*. In the case of FPLQ, there would be a great many coordinates sent to the LBS, and if any one of them were obvious fakes, the entire fake trajectory would fail.

You, Peng and Lee [15] refined the direct spatial shift technique. Instead of adding or subtracting an offset to (x,y) coordinates, they proposed rotating entire trajectories, including past and future points. They did not explain how they could obtain future points of a trajectory, however in vehicular contexts this is possible. Since BSMs are transmitted by all cars, a local vehicle could use other vehicles' past BSM transmissions to manufacture a new future fake/dummy trajectory.

This chapter presents random rotation of vehicle trajectories (RRVT). The approach was to examine one generic dummy-based scheme applied in a vehicular

context. The results were that, with modifications, the scheme may be useful in certain vehicular situations.

The foundation of this research is presented in [5], which attempted to minimize the number of dummies required for given levels of short term disclosure (*SD*), long-term disclosure (*LD*) and distance deviation (*dst*). [5] evaluated human-like trajectories of mobile users, as illustrated in Figure 4.1 (left). This research evaluated vehicle-like trajectories as illustrated in Figure 4.1 (right). The paper [5] proposed two dummy generation methods, random pattern scheme and rotation pattern scheme. The random pattern scheme would arbitrarily choose a starting point, ending point and points in between for dummy trajectories. The rotation pattern scheme would ensure overlap, extending the random pattern scheme by also arbitrarily choosing an intersection point and rotation angle for dummy trajectories. Recall: The more trajectories overlap, the lower the *LD*. See Figure 4.2.

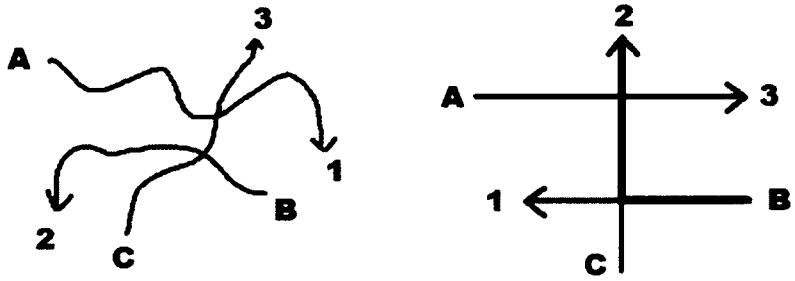


Figure 4.1: Trajectories. Human-like trajectories (left). Vehicle-like trajectories (right).

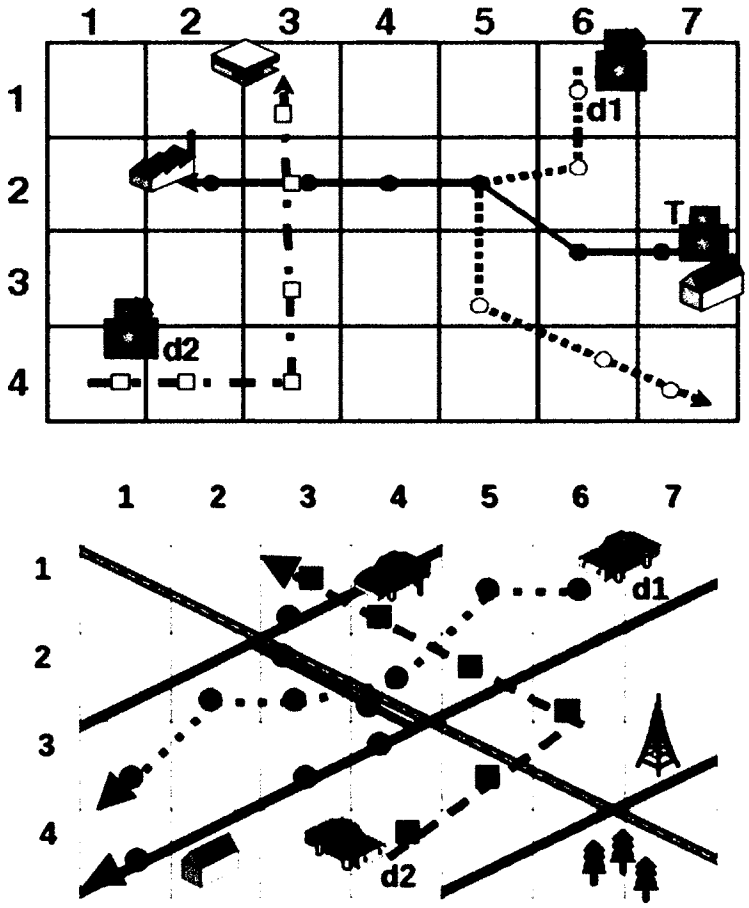


Figure 4.2: Mobility patterns. Human-like patterns (top) and vehicle-like patterns projected over roadways (bottom).

4.1 Demonstration

To illustrate the difference between human-like trajectories and vehicle-like trajectories consider Figure 4.2. While humans may roam freely and move relatively slowly, vehicles tend to move in more predictable patterns much more quickly. Figure 4.2 (top) shows how dummies *d1* and *d2* may be undetectable as dummies because their movement patterns are human-like.

Figure 4.2 (bottom) shows how dummies *d1* and *d2* are detectable as fakes in vehicular contexts. Dummy d1 is detectable because it does not follow a vehicle-like pattern. Dummy *d2* is detectable because, while it moves in a vehicle-like pattern, some of the positions are not on roadways, or it is easy to detect that it is impossible to move from one position to another by using a known roadway. Numerical data for the true user and dummies in Figure 4.2 (bottom) are presented in Table 4.1.

From equations (1), (2) and (3) we can compute *SD*, *LD* and *dst* for this user. See equations (4), (5) and (6) below.

$$SD = (1/6) * (1/3 + 1/3 + 1/2 + 1/3 + 1/3 + 1/3) = 0.3611 \quad (4.1)$$

$$LD = 1 / (5 + (3 - 2)) = 0.1667 \quad (4.2)$$

$$dst = (1/6) * (3.1 + 2.2 + 1.0 + 1.4 + 1.8 + 2.3) = 2.0 \quad (4.3)$$

The scenario above illustrates the random pattern scheme. To illustrate the implications of the rotation pattern scheme consider Figure 4.3. For human-like movement rotation angle may be chosen arbitrarily. For vehicle-like movement, because of often perpendicular roadways, it may be more advantageous to constrain dummy trajectories to rotations in increments of 90 degrees.

Table 4.1

Privacy Measurement of Random Pattern Scheme Trajectories

Time slot, *i* (*m* = 6)	1	2	3	4	5	6
True vehicle location	(3,1)	(3,2)	(4,2)	(4,3)	(3,3)	(1,4)
Dummy 1 location	(6,1)	(5,1)	(4,2)	(3,2)	(2,2)	(1,3)
Dummy 2 location	(4,4)	(3,3)	(6,2)	(5,2)	(4,1)	(3,1)
$\|Di\|$, number of unique locations	3	3	2	3	3	3
Distance (dist())	3.1	2.2	1.0	1.4	1.8	2.3

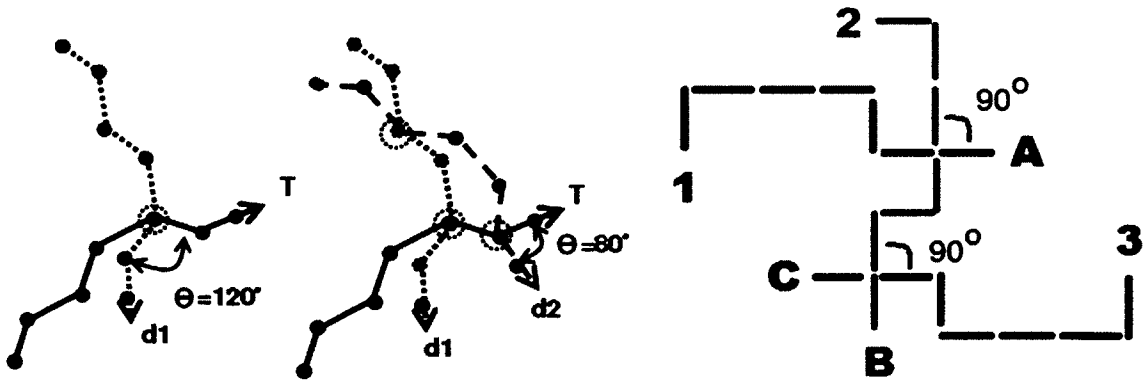


Figure 4.3: Rotation patterns. Human-like rotation patterns (left), and vehicle-like rotation patterns (right).

Restricting the rotation angle offers both advantages and disadvantages. Constraints may be considered disadvantageous because vehicles have fewer potential paths to choose from as they move, which degrades *SD*. However, fewer potential positions for vehicles implies more potential overlap, which may improve *LD*. Consider the data in Table 4.2. Compare the illustration in Figure 4.4 with Figure 4.2 (bottom). The former shows trajectories more overlapping positions. Since vehicles frequently transmit precise positions it is possible for a vehicle to construct realistic dummy trajectories using real or realistic data from other vehicles on the road.

From equations (1), (2) and (3), we can compute *SD*, *LD* and *dst* for this user. See equations (7), (8) and (9) below.

$$SD = (1/6) * (1/3 + 1/3 + 1/1 + 1/2 + 1/2 + 1/3) = 0.5278 \quad (4.4)$$

$$LD = 1 / (9 + (3 - 3)) = 0.1111 \quad (4.5)$$

$$dst = (1/6) * (3.8 + 2.6 + 0.0 + 1.0 + 3.0 + 3.6) = 2.0 \quad (4.6)$$

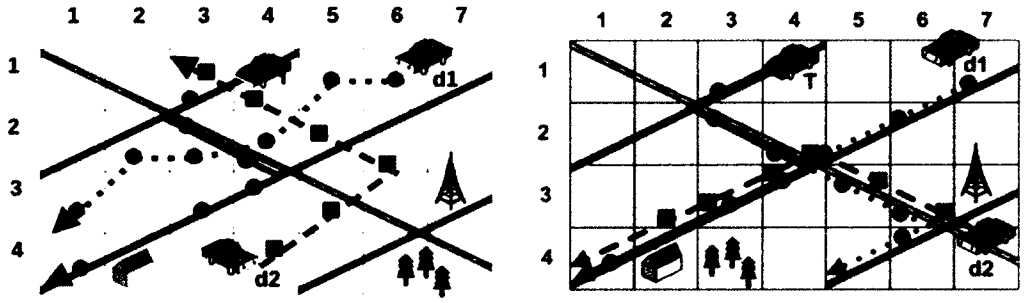


Figure 4.4 Modified rotation patterns. Human-like trajectories over roadways (left), and vehicle-like trajectories (right).

Table 4.2

Privacy Measurement of Random Pattern Scheme Trajectories

Time slot, i (m = 6)	1	2	3	4	5	6
True vehicle location	(3,1)	(3,2)	(4,2)	(4,3)	(3,3)	(1,4)
Dummy 1 location	(7,1)	(6,2)	(4,2)	(5,3)	(6,3)	(6,4)
Dummy 2 location	(6,3)	(5,3)	(4,2)	(4,3)	(3,3)	(2,3)
|Di|, number of unique locations	3	3	1	2	2	3
Distance (dist())	3.8	2.6	0.0	1.0	3.0	3.6

4.2 Simulation

We simulated scenarios similar to the illustration above, except we used 20 time slots and 5 to 25 dummies on a grid of 50x50 squares. We computed *SD*, *LD* and *dst* for each scenario for each of two conditions, one where roadways were restricted to exist only in squares which had one dimension evenly divisible by 10, the other with no such restriction. We ran each scenario nine times and recorded the run with the median number of trajectory intersections.

The charts in Figure 4.5 show how restricting locations to realistic road paths reduces the chance of long term disclosure, *LD* (Figure 4.5, right), while maintaining minimal effect on short term disclosure, *SD* (Figure 4.5, left).

4.3 Summary

Dummy locations which are not realistic in vehicular contexts may be detectable (map deanonymizable) as fakes because location coordinates can be cross-referenced and validated using maps. This problem can be resolved by restricting dummy vehicles to

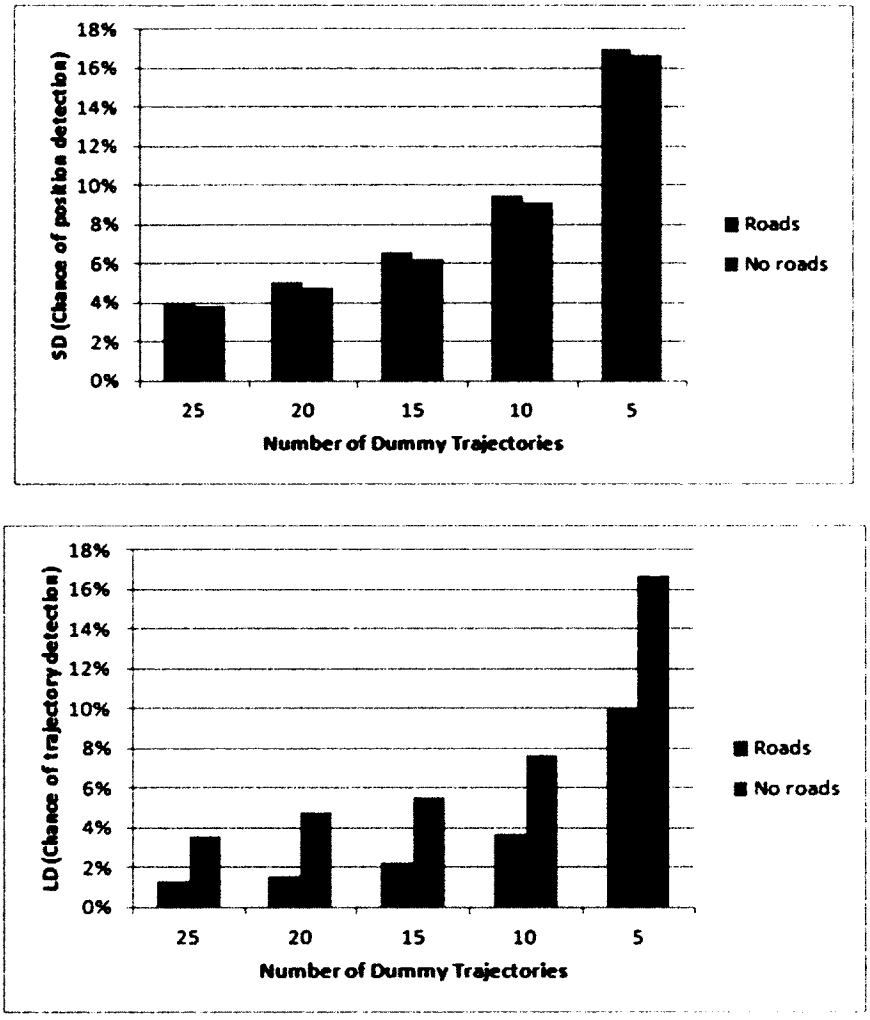


Figure 4.5: Effect of realistic roadway rotation on deanonymization. The chart at left shows short term disclosure. The chart at right shows long term disclosure. The left bars utilize human-like rotation patterns. The right bars utilize vehicle-like rotation patterns.

roadways. This reduces the number of locations and therefore increases the number of overlapping trajectories which improves *LD*. The method in [5] modified by a 90 degree rotation, instead of random rotation, is more realistic in vehicular context and better protects privacy because the chance of long term disclosure is reduced.

Two key areas of future work include evaluation of realistic distance deviation and frequency of LBS requests. The foundation paper in [5] and our paper present purely abstract quantities. See Table 4.3. In vehicular settings we can and should estimate the privacy protection using realistic distances which will likely depend on the precision of GPS devices used in on-board VANET components. Continuous precise location tracking also remains a problem even more challenging than the general vehicle location tracking problem. If LBS receives dozens, hundreds or even thousands of requests per hour the privacy of a vehicle may become more difficult to protect. Further, dummy trajectories may become so numerous that the resulting congestion on the LBS server due to unnecessary database queries may render the technique impractical.

Table 4.3

Simulation Data for Scenarios

(a) Roads restricted to every 10 squares in grid					
dummies	25	20	15	10	5
***rr* = 10**	10	10	10	10	10
SD	0.04052	0.05031	0.06547	0.09464	0.17
LD	0.01282	0.01538	0.02272	0.03703	0.1
intersects	26	22	14	8	2
dst	25.793	22.7686	26.5048	25.8764	20.6967

(b) No road restrictions (*rr*)					
dummies	25	20	15	10	5
***rr* = 0**	0	0	0	0	0
SD	0.03853	0.04761	0.06270	0.09136	0.16666
LD	0.03571	0.04761	0.05555	0.07692	0.16666
intersects	1	0	1	1	0
dst	23.8404	24.5140	25.6914	23.8121	31.6829

CHAPTER FIVE

ENDPOINT PROTECTION ZONE (EPZ)

The US DOT is expected to require all vehicles to transmit BSMs. If they require vehicles to transmit BSMs from the time the ignition system starts through the time the vehicle is turned off, and if the BSMs are transmitted every 100 ms all along the way as they are expected to be, then any eavesdropper with a sufficiently detailed map and sufficiently broad coverage of receivers could identify the driver and track the trajectory of any vehicle.

Consider that home owner information is publicly available digitally, and Google Maps API provides latitude and longitude information free of charge. When a vehicle ignition starts a vehicle, the latitude and longitude could be cross referenced with Google Maps API to find an address, then the address could be cross referenced with public records of home owners to identify the driver.

On one hand this could give law enforcement a *force multiplier*, the capability to tail suspects without expending the resources of an unmarked car. On the other hand it could also give drug dealers the ability to monitor suspected undercover cops, or their families, without requiring the crooks to be physically present at the location under surveillance. Even without their own network of receivers, legitimate administrators, malicious administrators, hackers or anyone with access to the mandatory TMS could achieve the same result.

In vehicular networks when map databases may be used to deanonymize user locations, we propose location based services, LBSs, be designed so that LBS users are

grouped by spatial location, into endpoint protection zones, EPZs. Users in the same EPZ would share login credentials (if login is required), and remain transmission-silent until outside of the EPZ, thus preventing an LBS administrator from knowing which particular user from the EPZ is active—even if the LBS administrator colludes with administrators of roadside units, RSUs. Simulations in this study used realistic vehicle traffic mobility models to measure improvements in privacy protection under varying EPZ sizes and vehicle densities.

This chapter is organized as follows. Section 5.1 discusses link layer threats and the special strengths and weaknesses of DSRC netowrks compared to other networks. Section 5.2 does the same, but for hight layers, focusing on applications that are likely to be part of an intelligent transportation system based on DSRC. Section 5.3 presents endpoint protection zones (EPZs) and how they might address collusion between LBS and RSU administrators (or hackers thereof). Section 5.4 presents the metrics sued in the simulations and section 5.5 presents the simulation setup and performance evaluation. Section 5.6 concludes the chapter.

Our contribution here is the EPZ model, a solution to the FPLQ problem under map deanonymization and collusion. We also measure the effectiveness of this solution.

5.1 Privacy Mechanisms, Link Layer (RSU)

Some researchers have proposed complex message anonymization and validation techniques which, while not concealing vehicles' precise locations, do conceal identities of vehicles. To ensure safety-related messages are valid, transmissions include public key certificates. DSRC standards enable message authenticity by providing standards for encryption through public key infrastructure (PKI). DSRC standards achieve identity

privacy by using temporary identifiers, i.e., temporary media access control addresses, temp-MACs, and temporary pseudo-identities, pseudoIDs, with their corresponding digital certificates. Permanent, real identifiers are not transmitted. Certificates are issued for pseudoIDs—thousands per vehicle—each valid for perhaps five or ten minutes during the course of a given year. [5]

Temporary identifiers provide only limited location privacy, however. On one hand, if a malicious system administrator, a tracker, were to mark a target using its temp-MAC or pseudoID, he would not be able to track a vehicle for very long because the identifier would periodically change, perhaps every five or ten minutes. On the other hand, if a single vehicle changed its identifier while all other nearby vehicles maintained their identifiers, then a tracker might determine a vehicle's current identifier is correlated to its prior one. Theoretically a tracker could track a vehicle for a period of time longer than the duration of the temporary identifier.

Researchers have proposed various solutions to the temporary identifier tracking problem. They have recommended groups of neighboring vehicles synchronize times (e.g., silent periods) and locations (e.g., mix zones) when and where they change pseudo-identities [6]. One subset of proposals describes a group model in which vehicles travel in clusters, all using the same group temporary identifier, and authenticating messages using the same group signature [7]. The group model has been shown to be effective in achieving anonymity in wireless/WSMP communications but it is less effective the lower the vehicle density, since anonymity level depends on the number of vehicles in each group. Some researchers suggest the group model is infeasible due to limitations of bandwidth and computation power, since pseudoID schemes create large certificate

revocation lists, CRLs, and associated checking costs, network overhead necessary to verify that certificates have not been revoked [8].

These solutions do not protect against map deanonymization.

5.2 Privacy Mechanisms, Higher Layers (LBS)

Motorists may be more vulnerable to location tracking when they access an LBS, such as the TMS. If login is required, the identity used to log in to the LBS may not be temporary, which may extend trackability time while logged in to the LBS. Further, transmissions and LBS queries may be correlated, deanonymized, using public databases, such as home address databases, which may diminish the identity privacy protection provided by temporary identifiers. Group signatures have been suggested as a solution to this problem, too [9], though we suggest group logins where the group is defined by an EPZ. Neither group signatures nor group login protect against map deanonymization, and neither defends against LBS/RSU collusion.

Some privacy models require a proxy server which acts as a privacy-protecting intermediary between vehicles and applications. Spatial cloaking schemes usually require a trusted server. The two basic problems with this technique are cost and trust. There are additional overhead and hardware costs, and there is the need to trust a server which can itself become a target or vulnerability.

Spatial cloaking is not always feasible in vehicular situations when LBSs require precise, continuous vehicle location information. While a single request can be obfuscated by a spatial cloak, a series of requests makes it ever more difficult to protect the link between the identity of a user and his location. If a user requests k,s-privacy in several snapshots, the snapshots might be assembled to identify which user is common to

all requests. Of course, if one end of the trajectory could be deanonymized, the user's identity might be linked with his location.

A few studies have explored the use of dummy events, i.e., counterfeit transmissions, in continuous, precise location situations. Instead of transmitting a spatially cloaked region in a single LBS request, a user would transmit multiple LBS requests, each containing a specific location, perhaps real, perhaps fake. Users would achieve location privacy by *k*-anonymity since LBS administrators could not tell which of *k* precise locations is genuine. The problem with this solution is under LBS/RSU collusion the LBS administrator could determine which locations were fake if the request used a false location. Even if the LBS user used real locations from real vehicles in its transmission range, the spatial range of the fakes would be limited by the vehicle's wireless communications range; that is, the decoy might be undetectable but it might be so close to the real vehicle that the location privacy achieved would be minimal.

These solutions do not protect against map deanonymization.

5.3 EPZ Model

To illustrate the value of the EPZ model, consider a hypothetical example, www.HomeOwnerMobile.com, a fictional LBS which displays on your car's dashboard the names of the owners of homes in view. It might also display the prices of the homes, if they are for sale, or the most recent price paid, if they are not for sale, assuming this is public information. Such a service would simplify the process of shopping for a home, but it would also require continuous precise location information from the vehicle making the request. What if the shopper does not want LBS administrators (or TMS hackers) to know his identity? How could the LBS be organized to prevent surveillance?

Suppose the LBS is constructed in such a way that LBS users who live near each other use the same login credentials. Suppose those users remain completely transmission-silent while driving in their specially designated area near their own home or workplace, i.e. their endpoint protection zone, EPZ. Then, if there are k users in the EPZ, each user achieves k-anonymity. The LBS administrator cannot identify which member of the EPZ is requesting the information. Moreover, since the users are from the same general location, location-specific advertising could still be effective.

5.3.1 EPZ Grids

In the EPZ model all LBS users in an EPZ protect each others' location privacy by using the same login credentials. The EPZ model divides a region, R, of width, W, and height, H, into grids of rectangles of width, w, and height, h. See Figure 5.1. Let V be the total number of vehicles in R. Let λ be the ratio of LBS users to V, so the number of LBS users in R is λV. Assuming a uniform distribution of random variable V in R, the expected anonymity set size for an LBS user is as follows.

$$E\{AS_{EPZ}\} = k = \lambda V wh/WH \tag{5.1}$$

5.3.2 Threat Model

Attackers can be categorized by the scope of their surveillance capabilities. If the attacker can observe the entire system of vehicles, we define him as a global attacker, even though the scope of the system may only include a single municipality. If the attacker has access only to a subset of the system, such as the communications range of an RSU, we refer to the attacker as local. Attackers can also be categorized by their intent, passive or active, i.e. whether they intend merely to monitor targets or whether they also intend to mislead or otherwise influence targets.

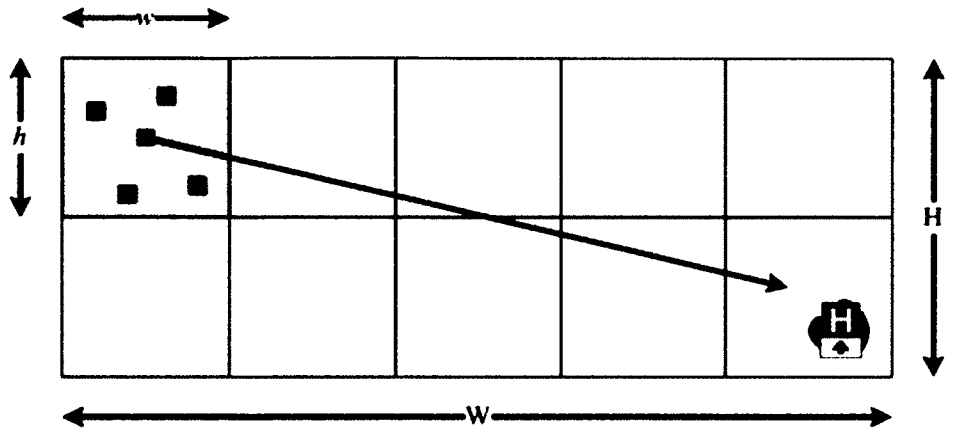


Figure 5.1 EPZ Grid. The proposed model divides regions into square sections, called endpoint protection zones, EPZs.

This chapter assumes a global passive adversary at the LBS colluding with a local passive adversary at the RSU nearest the vehicle under surveillance. This global-local adversary may be an "insider" with legitimate authority to monitor these systems, or the attacker may have acquired/hacked such access illegally.

We assume the adversary wishes to determine the location of a target vehicle, and that the adversary may have already linked the target's LBS userID with a pseudo-identity of a vehicle within range of an RSU or another vehicle. Table 5.1 outlines potential LBS administrator attack scenarios.

5.3.3 Limitations of the Model

For each LBS user driving inside his EPZ, this model prevents sending safety and traffic management data, reducing some functionality of the system. The model also precludes infotainment while in one's own EPZ. For other vehicles, the model degrades safety and traffic management functionality to the extent vehicles are operating in their

Table 5.1

Location Privacy Attacks by LBS Administrator

LBS User Action	**LBS Administrator Attack**
Transmit precise home location	Identify user with deanonymization*
Obfuscate position with spatial cloaking	Isolate trajectory from snapshots, then identify user with deanonymization*
Do not transmit to LBS inside EPZ, but always transmit safety messages	Isolate trajectory from snapshots, then collude with RSU to correlate paths, then identify user with deanonymization*
Remain transmission silent within EPZ	Cannot deanonymize* (no identifiable endpoint)

* LBS administrator uses endpoint (origination or termination) to deanonymize

own EPZ. We envision EPZs to be as small as possible to mitigate these limitations. Also, not all LBS users will care about privacy. This model protects only those users who do. The model does not defend against license plate readers, mobile phone monitoring, roadside cameras, and physical surveillance.

5.4 Metrics

In the VANET security literature, privacy is often equated to anonymity. Even the IEEE 1609.2 (2013) security standard [4] itself does so, saying, "Anonymity—meaning the ability of private drivers to maintain a certain amount of privacy—is a core goal of the system." A frequently used metric is k-anonymity, though others include l-diversity [10] t-closeness [11], and ε-differential privacy [12]. This paper confines itself to the

concept of k-anonymity, or anonymity set size, defined in the Introduction of this document. In this paper we measure privacy by anonymity set size, entropy of the anonymity set size and tracking probability. For a discussion of this topic, see [13].

This study used three metrics, anonymity set size, entropy of anonymity set size and tracking probability.

5.5 Simulation

To prepare the simulation we used realistic vehicle mobility models and estimated privacy levels using custom simulation software.

5.5.1 Mobility Patterns

Computer simulations do not always represent vehicle traffic flows accurately. Harri, et al. [15] suggest that minimum requirements for realistic simulations include techniques for intersection management, lane changing and car following. Several systems offer these features, including Generic Mobility Simulation Framework, GMSF [16]. We used Multi-agent Microscopic Traffic Simulator, MMTS, trace files linked from the GMSF website [17] and provided at the Laboratory for Software Technology website [18], specifically City, Urban and Rural. All three models contain records of time-stamps, vehicle-ids, x-coordinates, y-coordinates within a 3000m x 3000m (9 million square meters) grid. Each model starts with a different number of vehicles, v. City starts with v=897. Urban starts with v=488. Rural starts with v=110. Vehicles enter and leave the system at roughly the same rate, so the number of vehicles in the model at any given time is not always precisely the same as the number at the start.

Road topologies in some mobility models, such as the Freeway model (a straight road with perhaps several lanes) and the Manhattan model (a grid of horizontal and

vertical roads), vehicle density per linear meter can be out of sync with the vehicle density per square meter, especially when compared with more realistic road topologies. For example, for 900 vehicles in a 3000x3000 meter grid, the Freeway model might have a linear density of 0.3 v/m, 900 vehicles divided by 3000 meters, and a square density of 0.0001 v/m2, 900 vehicles divided by 9 million square meters. The Manhattan model would have a linear density of 0.004839 v/m, 900 vehicles divided by 186,000 meters, but the same square density as the Freeway model. In other words, the linear density of the Manhattan model is 1.6% that of Freeway model given the same square density.

Our simulation does not have this problem because the linear distances covered by the road topologies are similar: City, 14,783 meters; Urban, 13,955 meters; and Rural, 10,175 meters. The areas covered are identical, so the mobility models we used provide both realistic traffic flows and comparable coverage distances and areas.

5.5.2 Metrics Computations

We wrote a program which read MMTS mobility model files, city, urban and rural. Each mobility model has a different vehicle density in the same size region. The program divided each 3000m x 3000m region into square EPZs, ranging from 300m x 300m (100 EPZs) to 1500m x 1500m (4 EPZs). For each mobility model the program computed the metrics, $|AS|$, $H(|AS|)$ and Pt. All simulations covered vehicle movements over a time period of 2000s, or 33.3 minutes.

5.5.3 Performance Evaluation

The EPZ model was effective only in sufficiently high density areas. When vehicle density, and therefore LBS user density, was low, and EPZ sizes were small, then

anonymity set sizes approached 1.0 and tracking probabilities approached 100%, which represents the poorest possible privacy protection under our metrics.

Figures 5.2, 5.3 and 5.4 show results when λ=10%. 5.5, 5.6 and 5.7 show results when λ=20%. Doubling λ doubled $|AS|$ and $H(|AS|)$, but at low densities it more than halved Pt.

Anonymity set size. Average anonymity set size for the high-density city model ranged from 5.26 (100 EPZs) to 50 (4 EPZs); for the medium-density urban model, from 4.84 to 30.25; for the low-density rural model, from 1.0 to 1.75. In other words, an attacker would have roughly a 1/5 to 1/50 chance of identifying the target vehicle in the city model; a 1/5 to 1/30 chance in the urban model; but a 1/1 to 1/2 chance in the rural model. Low density settings would require a different method, or EPZs larger in size than the sizes we tested.

Entropy of the anonymity set size. Average entropy ranged from 2.39 to 5.64 for city; 2.27 to 4.92 for urban; and 0 to 0.81 for rural. Again, low density settings would require a different method, or EPZs larger in size than the sizes we tested.

Tracking probability. Average tracking probability ranged from 7% to 0% for the city model; 7% to 0% for the urban model; and 100% to 14% for the rural model. The model is of little use in areas of low density given the EPZ sizes we tested.

In practice the model may need to be modified so that the density of LBS users determines EPZ size, not the other way around. Further experimentation may determine the optimal breakeven EPZ size given the tradeoff between privacy protection and service degradation.

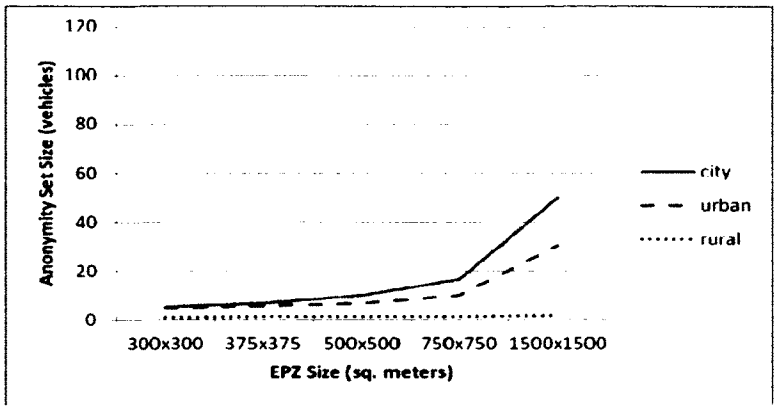


Figure 5.2: Average anonymity set size by EPZ size, 10% LBS users

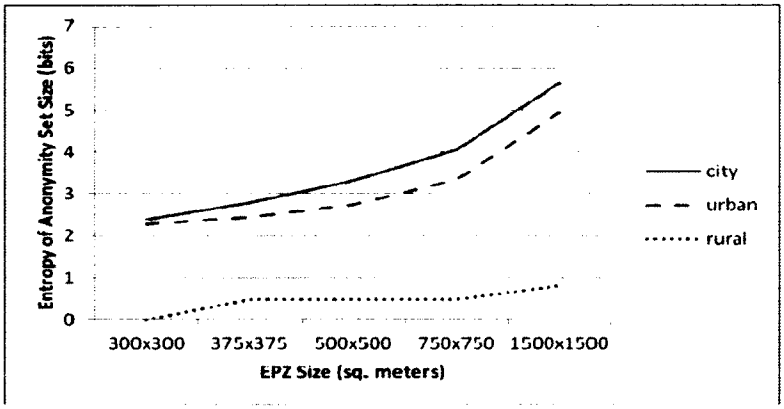


Figure 5.3: Entropy of average anonymity set size vs. EPZ size, 10% LBS users

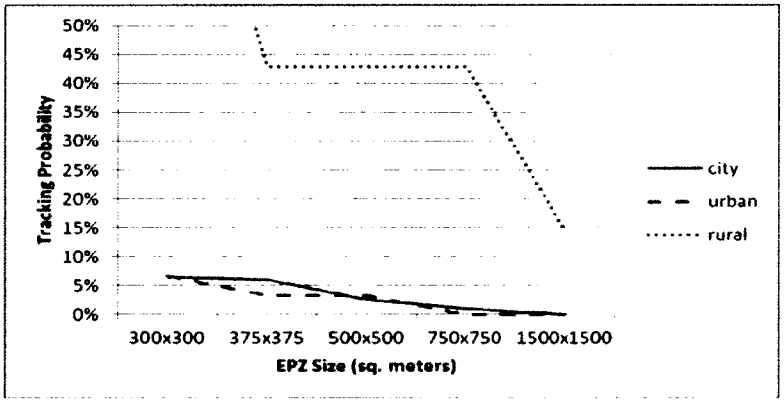


Figure 5.4: Tracking probability vs. EPZ size, 10% LBS users.

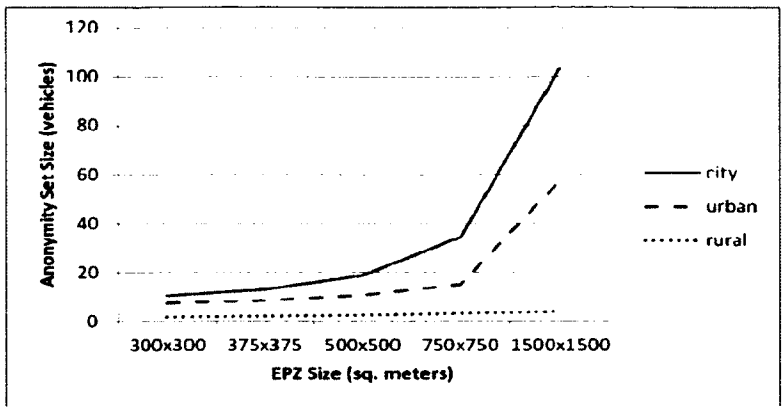


Figure 5.5: Average anonymity set size vs. EPZ size, 20% LBS users

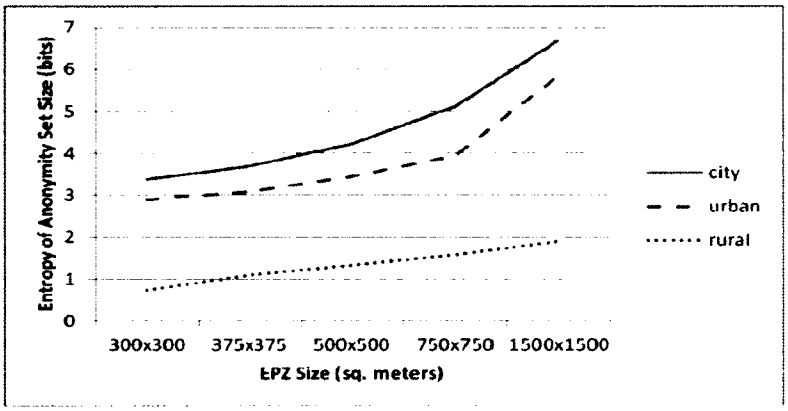


Figure 5.6: Entropy of average anonymity set size vs. EPZ size, 20% LBS users

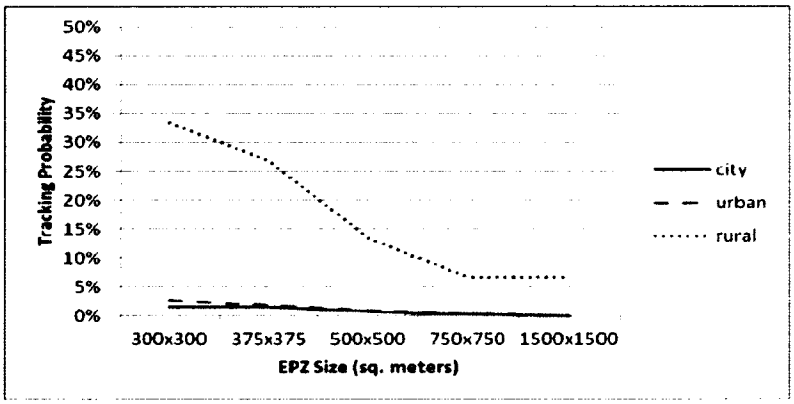


Figure 5.7: Tracking probability vs. EPZ size, 20% LBS users

In computing the total number of vehicles, V, we ignored vehicles whose trajectories originated at the edge of the region. Vehicles whose trajectories originated on the edge were assumed to belong to EPZs located outside of the region. More details and the program code for this simulation can be obtained from www.vanetprivacy.com.

5.6 Summary

Endpoint protection zones, EPZs, protect vehicle location privacy from deanonymization. If LBS administrators can correlate origin and destination points with home and work addresses, they can link identity and location of vehicle owners. This is not possible if vehicles remain transmission-silent in their respective EPZs.

EPZs protect vehicle location privacy from collusion between LBS administrators and RSU administrators. Even if LBS administrators can verify with RSUs the locations of transmissions at their points of origination, they cannot be certain which vehicle from the EPZ is making the request unless they have additional information beyond the scope of this study.

The effectiveness of EPZs depends upon density, multiple LBS users originating from each EPZ. In sparsely populated areas, the EPZ model may be ineffective. One workaround might be encouraging local friends and family to use LBS. Another workaround might be to increase EPZ sizes in sparsely populated areas. The practicality of this is a subject of future study. Perhaps the most important finding of this paper is that a small increase in LBS users in a sparsely populated area, which may yield only a proportional effect in anonymity set size, may have a much greater effect on tracking probability.

An interesting property of the EPZ model is that it provides protection even if only one LBS user from an EPZ is active outside that EPZ. The LBS administrator cannot know which LBS user is active. This implies that LBS users especially concerned about privacy could register under multiple false identities, or have friends and family who do not use the LBS register under real identities. In this way a single person could achieve k-anonymity > 1.0.

EPZs do not protect against many forms of surveillance, such as license plate readers, mobile phone monitors, roadside cameras or physical surveillance. However, against collusion and deanonymization attacks, EPZs may be a useful tool in high density areas to protect vehicular location privacy.

CHAPTER SIX

PRIVACY BY DECOY (PBD)

The weakness of the RRVT model is it only works over short distances and it is deanonymizable at the RSU level. The weakness of the EPZ model is it only protects endpoints; if a car is *marked* in transit it remains vulnerable from the point at which it was marked until it conceals itself in the terminal EPZ. The next logical advancement in the protection of location privacy against eavesdroppers is a solution which overcomes these weaknesses.

This chapter introduces PARROTS, Position Altered Requests Relayed Over Time and Space, a privacy protocol which protects LBS users' location information from LBS administrators even (1) when the LBS requires continuous precise location data in a vehicular ad hoc network, (2) when LBS administrators collude with administrators of vehicular wireless access points (a.k.a. roadside units, or RSUs), and (3) when precise location data can be deanonymized using map databases linking vehicle positions with vehicle owners' home/work addresses and geographic coordinates. Further, this protection remains in force over greater distances.

The PARROTS protocol is also called Privacy by Decoy (PBD), to distinguish it from a naïve dummy event model. Recall: A *dummy* is a fake location sent by the vehicle desiring location privacy. A *decoy* is a real location sent by a vehicle other than the vehicle desiring location privacy.

PBD as described in this chapter is an extension of EPZ. Recall: The analysis of the EPZ model [27] showed that location based services, LBSs, be designed so that LBS

users be grouped by spatial location, into endpoint protection zones, EPZs. Users in the same EPZ would share login credentials, and remain transmission-silent until outside of the EPZ, thus preventing an LBS administrator from knowing which particular user from the EPZ is active—even if the LBS administrator colluded with administrators of roadside units, RSUs [6]. This section further proposes LBSs be designed so that users relay dummy/false queries through non user vehicles to camouflage true locations.

Our main contributions are (1) a new way of looking at vehicular location privacy, called privacy-by-decoy, PBD, (2) a new PBD model, and (3) an examination of EPZ alone compared to EPZ with PBD. The representative PBD model is called Position Altered Requests Relayed Over Time and Space, or PARROTS. This model is meant to apply in vehicular contexts, but it could apply in other contexts as well.

Moreover, unlike any other model of which we are aware, PARROTS functions even when the defender is inactive. Decoys provide cover by relaying LBS requests even when the target of an attack turns off her car. To underscore this point, under other models an attacker perhaps may not locate the target, but at least he will know the target is active. Under PARROTS, the attacker cannot be sure that an attacker is active even if he is receiving an active signal because the signal could be coming from a decoy.

Still further, unlike any other model of which we are aware, PARROTS functions even in locations from which the target could not possibly transmit. Decoys provide cover by relaying LBS requests perhaps far, far away from the actual location of the target. To underscore this point, under other models targets transmit their own fake queries, so an attacker can determine that a too distant decoy message could not possibly originate from within a certain geographical area. Under PARROTS, the decoy could

transmit a request from a location hundreds of miles from the target and the attacker could not determine if the signal was a fake or if the target just took a long trip.

Besides the EPZ model we are aware of no study to date which has examined the map deanonymization of endpoints in VANETs under LBS/RSU collusion. This section extends the prior solution to this problem offered by EPZ.

6.1 VANET System Model

Wireless vehicle-to-vehicle, V2V, communication takes place on an as-needed basis, as does vehicle-to-infrastructure, V2I, communication. In the latter case, a wireless access point called a roadside unit, RSU, relays network packets from vehicles to systems, such as LBSs, connected to the wired infrastructure. Privacy is handled differently in V2V and V2I communications.

6.2 Threat Model

Attackers can be categorized by the scope of their surveillance capabilities. If the attacker can observe the entire system of vehicles, this is a global attacker, even though the scope of the system may only include a single municipality. If the attacker has access only to a subset of the system, such as the communications range of an RSU, this attacker is local. Attackers can also be categorized by their intent, passive or active, i.e. whether they intend merely to monitor targets or whether they also intend to mislead or otherwise influence targets in some way.

This section assumes a global passive adversary at the LBS colluding with a local passive adversary at the RSU nearest the vehicle(s) under surveillance. This global-local adversary may be an "insider" with legitimate authority to monitor these systems, or the attacker may have acquired/hacked such access illegally. This paper assumes the

adversary wishes to determine the location of a target vehicle, and that the adversary may have already linked the target with a certain pseudo-identity of a vehicle.

6.3 EPZ Equation

The EPZ model [6] divides regions into grids of rectangles of width, w, and height, h. Let V be the total number of vehicles in a region, R, of area, A. Let λ be the ratio of LBS users in the same region, so the number of LBS users in R is λV. The expected anonymity set size for an LBS follows.

$$E\{ AS_{EPZ} \} = \lambda V w h / A \tag{6.1}$$

In the EPZ model all LBS users in an EPZ protect each others' location privacy by using the same group login credentials. This is enforced by the LBS when the user registers with the service; the EPZ is calculated at the point of registration after which the location information is discarded.

6.4 PARROTS Protocol

The EPZ Model can be enhanced by using decoys. Consider a regime in which helper vehicles, called *parrots*, relay LBS requests on behalf of an LBS user in a vehicle desiring privacy, called a *pirate*. The pirate would transmit LBS requests normally. Parrots would transmit LBS requests on behalf of the pirate, using the pirate's login credentials but the parrots' locations. Parrots' locations could not be identified as fakes because they are real locations in real current traffic conditions. Parrots could mimic pirates over great temporal and spatial range.

In order for this scheme to work LBS systems must permit login credentials to be sent encrypted, but vehicle locations unencrypted. After all, if the LBS users' credentials are not encrypted a malicious parrot could misuse the pirate's credentials. And if the

location is encrypted the parrot would not be able to perform the encryption since the parrot would not possess the pirate's private key.

Under this system there is no leakage of any information from the pirate to the parrot save the destination address of the LBS. Encryption prevents parrots from reading pirate's login information, query information or responses from LBSs since LBSs encrypt replies with the pirates' public keys.

Under this system the LBS cannot determine which vehicles are pirates and which are parrots. The login credentials are authentic, and the locations are real. Location privacy is achieved by k anonymity if a pirate has k-1 parrots. If LBSs permit group logins then even greater anonymity is accomplished.

By relaying each others' dummy queries, vehicles can protect each other from surveillance by LBS administrators – even if those administrators collude with administrators of other system infrastructure components such as RSUs. This protection can be accomplished if LBS software accepts authorizations using encrypted group userIDs, passwords and service requests, along with unencrypted locations.

In a traditional online system, such as Foursquare, used at a desktop or laptop computer or on a smart phone, a user may log in manually with a userID and password. Then, either the user may manually enter a location query, or the application may infer the location from contextual information such as the IP address or GPS information of the client computer. Using group signatures, authorization could be accomplished on a group basis. A traditional service permits users to log in from any computer. A secure session, such as occurs under HTTPS, would be authenticated using SSL certificates of the

computers, not of the LBS userIDs. So it is possible to allow multiple secure logins of LBS users. Anonymity sets may be formed as in Figure 6.1.

The PARROTS protocol suggests that LBSs should operate like traditional services, except the login information should be enclosed in an encrypted message, and the location information should be unencrypted. A vehicle which wishes to be parroted, i.e. a pirate, would send an encrypted LBS authentication message to a vehicle which is willing to relay the decoy messages, i.e. a parrot. The parrot would relay to the LBS the pirate's encrypted message appended with the parrot's location. The pirate would send to the LBS the pirate's encrypted message appended with the pirate's location. The result is

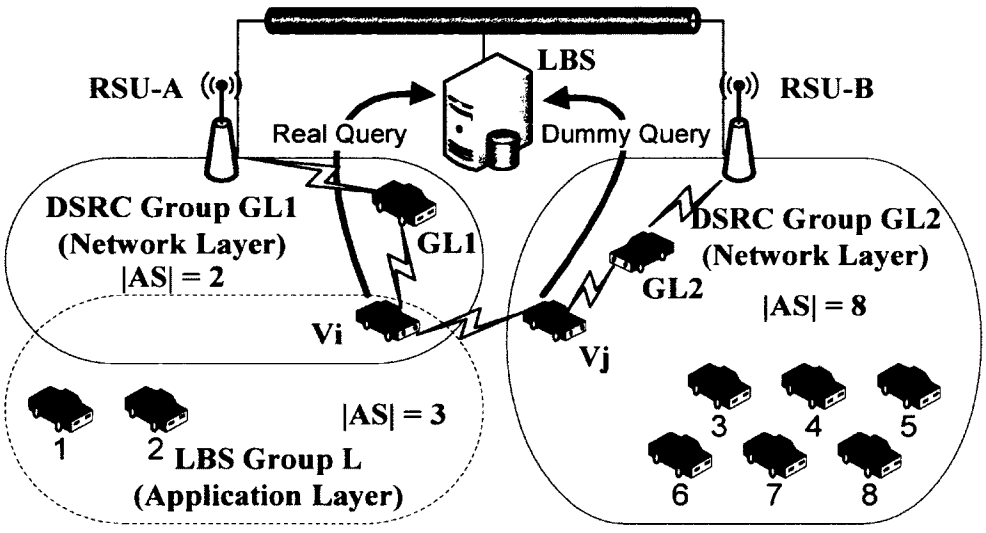


Figure 6.1: LBS Anonymity Sets. LBS users form groups to protect themselves from potentially snooping LBS administrators. Individual LBS users may communicate with other vehicles which agree to relay the LBS user's dummy queries, increasing the anonymity set size, $|AS|$, of the entire LBS group.

that the LBS administrator does not know which location is the location of the pirate – even if he can verify through IP traceback which vehicle is the source of each message.

Figure 6.2 shows how vehicle V_j could be contacted by vehicle V_i while in communication range. V_j would send no genuine queries of its own to the LBS, but rather it would send fake queries on behalf of V_i. Normally, the LBS would think V_i's request was from one of the members of LBS Group L, but in reality one query is coming from V_j, posing as a member of LBS Group L. If the LBS administrator performs an IP traceback to RSU B, he will confirm the vehicle sending the transmission is in fact at the location queried, $[X(V_j), Y(V_j)]$. So, with parroting, the anonymity set size increases by 1 for every member of Group L. This is shown in Figure 6.1.

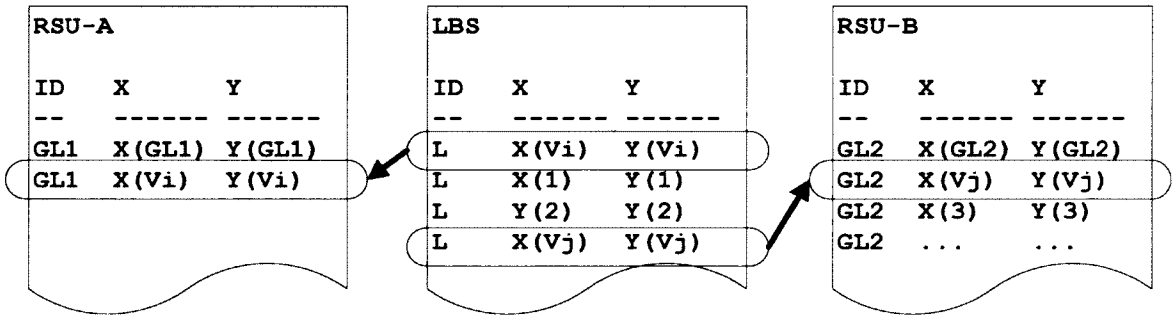


Figure 6.2: LBS-RSU Collusion. The reports above provide the information available to administrators of RSU-A, LBS and RSU-B, respectively. Notice that the |AS| of group L rises from 3 to 4 because of Vj, even though Vj is not part of LBS. If LBS administrator performs IP traceback, he finds query originated from the location queried.

6.4.1 Definitions and Assumptions

PARROTS depends on PKI. Each party in a secure communication has a public and private key. Define CA as the certificate authority which issues identities and certificates, including pseudo identities, or pseudoIDs, for vehicles. Let V_i indicate a vehicle with pseudoID *i*. Let V_j indicate a vehicle with pseudoID *j*. Assume pseudoIDs are valid for short periods of time, say five to ten minutes.

Let RSU refer to a roadside unit, a wireless access point for vehicles to connect to the wired infrastructure. Let LBS indicate a location based service. Let POI mean "point of interest," such as a restaurant or gas station. Let "request" be defined as a message asking for information, such that an LBS request would be a message asking for information from the LBS. Let U_i indicate the identity of LBS user in Vehicle *i*. Assume U_i is an email address, so it can have an associated public key, indicated by a trailing plus sign, U_i^+. Define $Cert(U_i)$ as the digital certificate binding U_i's email address and public key such that $Cert(U_i) = CA^-(U_i, U_i^+)$.

Let "pirate" be defined as the vehicle of an LBS user who wishes to be mimicked. Let "parrot" be defined as a vehicle, not necessarily linked to any LBS user, willing to mimic the pirate. If V_i wishes to be mimicked, then V_i would be a pirate. If V_j is willing to mimic V_i, then V_j would be a parrot.

Assume LBS users have userID/password combinations which do not expire as quickly as vehicular pseudoIDs and temporary MAC addresses. Further assume LBS users may log in to the LBS from any computer/vehicle.

6.4.2 PARROTS Equations

The PARROTS model depends on EPZs. Recall equation (6.1). Now let ρ be the ratio of potential parrots, so the number of potential parrots in R is ρV. Note that the set of LBS users and the set of potential parrots are disjoint sets. Now let φ be the ratio of LBS users who desire privacy. Note if all LBS users desired privacy, φ=1. The expected anonymity set size for an LBS user under the EPZ model with PARROTS if the LBS enforces single-user login is as follows.

$$E\{AS_{EPZ}\text{-}pi\} = 1 + \rho / \varphi \lambda \tag{6.2}$$

The expected anonymity set size for an LBS user under the EPZ model with PARROTS, if the LBS allows group logins, is as follows.

$$E\{AS_{EPZ}\text{-}pg\} = (\lambda + \rho)\, wh/A \tag{6.3}$$

6.5 Simulation

A simulation system was written using realistic vehicle mobility models conforming to standards in [22] which suggest that minimum requirements for realistic simulations include techniques for intersection management, lane changing and car following. Several systems offer these features, including Generic Mobility Simulation Framework, GMSF [23], the website for which offers Multi-agent Microscopic Traffic Simulator, MMTS, trace files linked from the GMSF website [24] and provided at the Laboratory for Software Technology website [25], specifically City, Urban and Rural. All three models contain records of time stamps, vehicle ids, x-coordinates, y-coordinates within a 3000x3000 meters (9 million square meters) grid.

Each model starts with a different number of vehicles, v. City starts with v=897. Urban starts with v=488. Rural starts with v=110. Vehicles enter and leave the system at

roughly the same rate, so the number of vehicles in the model at any given time is not always precisely the same as the number at the start.

Road topologies in some mobility models, such as the Freeway model (a straight road with perhaps several lanes) and the Manhattan model (a grid of horizontal and vertical roads), suffer from vehicle density challenges. Vehicle density per linear meter can be out of sync with the vehicle density per square meter, especially when compared with more realistic road topologies. For example, for 900 vehicles in a 3000x3000 meter grid, the Freeway model might have a linear density of 0.3 v/m, 900 vehicles divided by 3000 meters, and a square density of 0.0001 v/m2, 900 vehicles divided by 9 million square meters. The Manhattan model would have a linear density of 0.004839 v/m, 900 vehicles divided by 186,000 meters, but the same square density as the Freeway model. In other words, the linear density of the Manhattan model is 1.6% that of Freeway model given the same square density. The simulation upon which this paper is based does not suffer nearly as severely from this problem because the linear distances covered by the road topologies are similar: City, 14,783 meters; Urban, 13,955 meters; and Rural, 10,175 meters. The areas covered are identical, so the mobility models provide both realistic traffic flows and comparable coverage distances and areas.

The simulation software read the mobility model file and for each mobility model and computed the traditional metrics, ASi, Hi and Pti,. All simulations covered a time of 2000s, or 33.3 minutes. The software divided a 3000m x 3000m region into square EPZs, ranging from 1500m x 1500m (4 EPZs) to 300m x 300m (100 EPZs). 6.6 Analysis

Figure 6.2 summarizes the results of the simulation. It compares AS_i, H_i and Pt_i for two models: EPZ alone and EPZ with PARROTS, group login. The software

simulated using 10% LBS users (λ=0.10) and 10% potential parrots (ρ=0.10 and φ=1.00). In computing V, the software ignored vehicles whose trajectories originated at the edge of the region. Vehicles whose trajectories originated on the edge were assumed to belong to EPZs located outside of the region.

6.6.1 EPZ Alone

Simulation showed that the EPZ model is effective to the extent that multiple LBS users have endpoints in EPZs. When vehicle density, and therefore LBS user density, is low, and EPZ sizes are small, then anonymity set sizes approach 1 and tracking probabilities approach 100%, which represents the poorest possible privacy protection under the privacy metrics. EPZ works, but only with sufficient vehicle and LBS user densities.

6.6.2 EPZ with PARROTS, Group Login

The PARROTS model with group login performed best in tests. The effect of EPZ with group login, combined with PARROTS, produced anonymity set sizes close to the theoretical value, and showed visible improvement in tracking probabilities.

6.6.3 EPZ with PARROTS, Individual Login

The PARROTS model with individual login performed worse than the EPZ model alone, except in low density situations PARROTS demonstrated equivalent results. The reason for this is because PARROTS' performance depends not on the density of LBS users but on the ratio of potential parrots to pirates. In all cases, regardless of vehicle density, anonymity set size was near 2, the theoretical value.

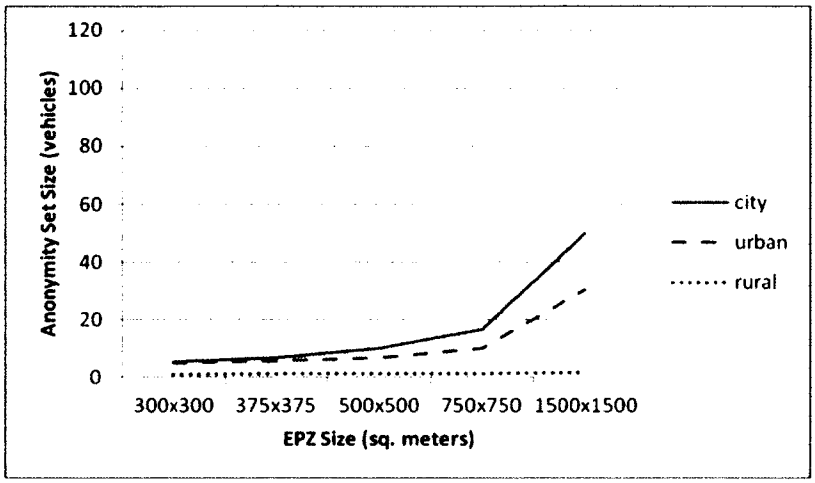
6.7 Summary

This chapter presented the concept of relayed dummy events as a privacy defense in FPLQ conditions in a VANET system under DSRC/WAVE. EPZ requires no additional network transmissions to establish privacy levels, so there is no bandwidth tradeoff for implementation. PARROTS requires a recruitment phase and multiple duplicate transmissions by parroting vehicles. This study measured the privacy performance, not the network efficiency, of the models. In future work we hope to run simulations to determine an optimal balance.

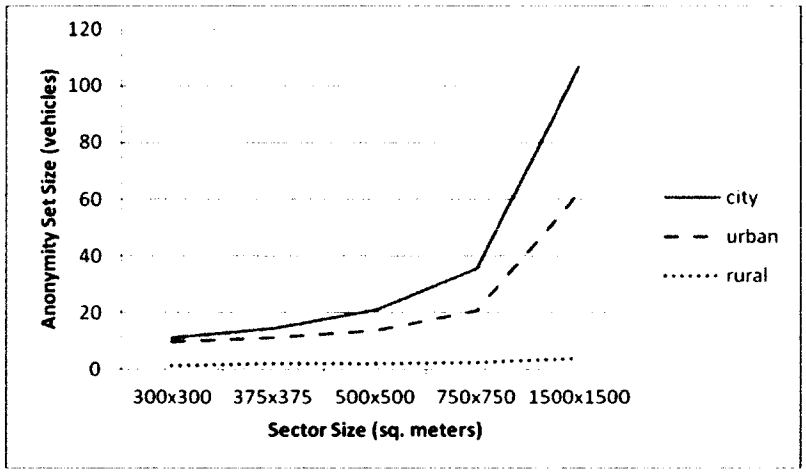
The EPZ and PARROTS protocols address conditions, (1) when the LBS requires continuous precise location data in a vehicular ad hoc network, (2) when the LBS administrator colludes with administrators of vehicular wireless access points (a.k.a. roadside units, or RSUs), and (3) when precise location data can be deanonymized using map databases linking vehicle positions with vehicle owners' home/work addresses and geographic coordinates. Simulations using realistic vehicle traffic mobility models showed PARROTS increased average privacy levels in high vehicle density conditions when deployed in combination with EPZs, and increased them even more effectively in low vehicle density conditions.

Neither PARROTS nor EPZs protect against many forms of surveillance, such as license plate readers, mobile phone monitors, roadside cameras or physical surveillance. However, against collusion and deanonymization attacks EPZs may be a useful tool to protect vehicular location privacy. PARROTS enhances the effectiveness of EPZs. Other vehicle location privacy methods which use decoys for protection would transmit dummy events while the real LBS querier is active. The PARROTS model is unusual in that

parrots can make requests of an LBS on behalf of pirates even when pirates are inactive. Under the specific conditions presented in this paper, it has the further benefit of enabling undetectable decoys over a spatial range broader than the communications range of the LBS querier. See Figure 6.3, parts (a) through (f), for graphical representation of results.

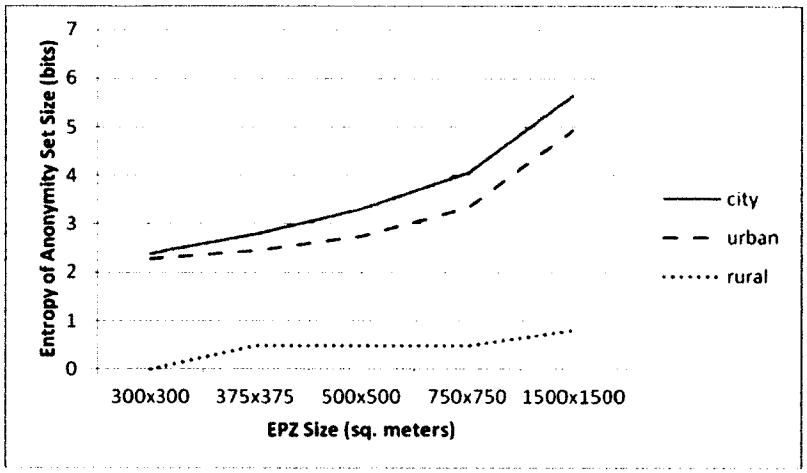


(a)

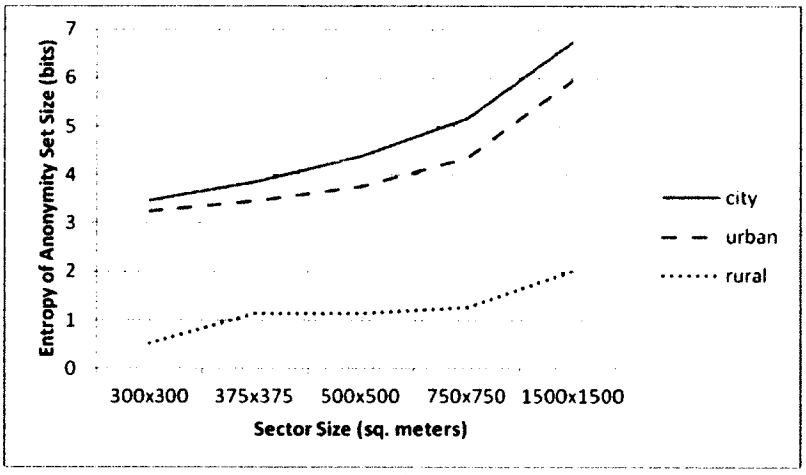


(b)

Figure 6.3: Comparison of EPZ and PBD Models. (a) EPZ alone, anonymity set size, (b) EPZ with PARROTS, group login, anonymity set size

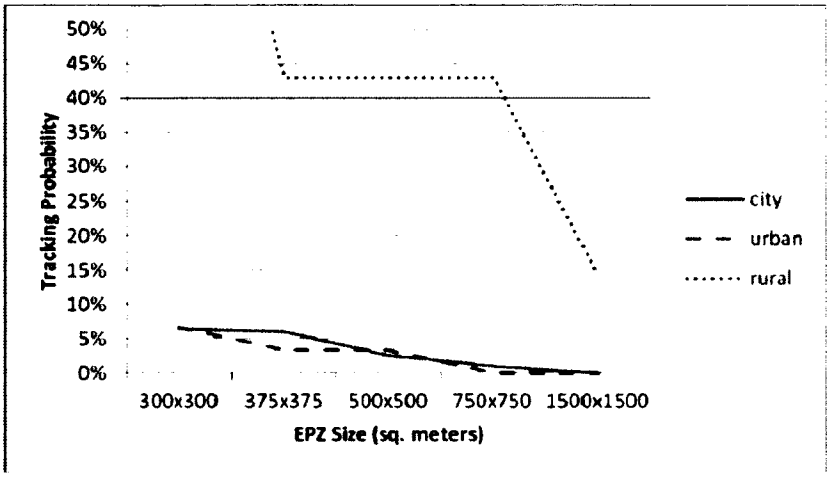


(c)

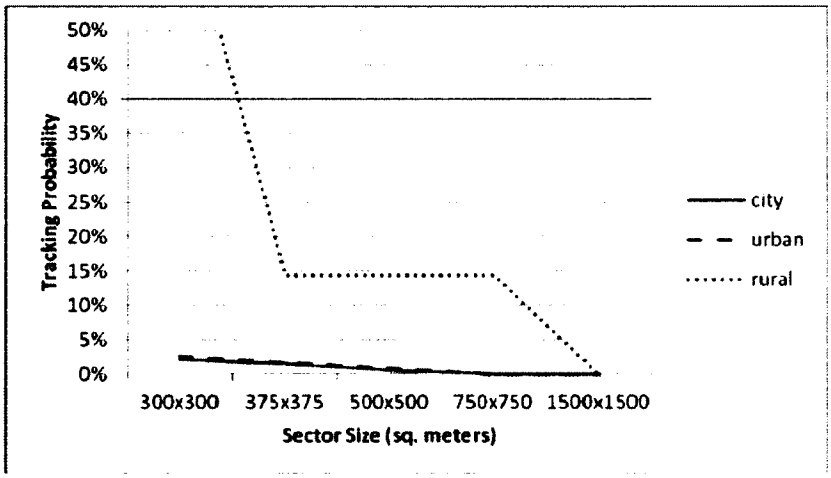


(d)

Figure 6.3: Continued... (c) EPZ alone, entropy of anonymity set size, (d) EPZ with PARROTS, group login, entropy of anonymity set size



(e)



(f)

Figure 6.3: Continued... (e) EPZ alone, tracking probability, (f) EPZ with PARROTS, group login, tracking probability

CHAPTER SEVEN

SAFETY-SILENCE TRADEOFF EQUATION (SSTE)

Automobiles and other vehicles will soon include network equipment, effectively turning roadways into moving communication systems. One of the primary objectives of such systems is safety. As has been discussed above, in future VANETs BSMs would be transmitted by all vehicles every 100ms in order to help prevent inter-vehicle crashes. At first not all vehicles would contain the hardware and software necessary to transmit BSMs; there would be an interval when only a percentage of vehicles would transmit. Further, even after all vehicles install the equipment, some privacy researchers recommend silent periods, spans of time during which vehicles deliberately cease transmissions.

Whether due to lack of equipment or due to privacy protocols, silent periods would defeat safety provided by BSMs. This chapter quantifies this tradeoff, presenting the *Safety-Silence Tradeoff Equation*, and showing an inverse exponential relationship between the proportion of vehicles transmitting BSMs and the proportion of potential collisions between vehicles unprotected by BSMs.

VANET equipment is expected to take time to roll out. To prevent a crash between two cars, the equipment must be operational in both cars. To obtain full safety value from a VANET requires equipment to be operational in all cars. It will take time to achieve near-full implementation. To our knowledge no study has been done to compute the cost of delaying such a rollout.

Further, VANETs raise privacy concerns because they would transmit data which may also be used for unwanted surveillance. Researchers have proposed protocols to protect against surveillance, but many protocols require periodic silent periods, spans of time when vehicles cease transmissions, thereby disabling VANET safety benefits. To our knowledge no study has been done to compute the cost of implementing these silent periods.

The privacy issue is not moot because of E911. It is possible for motorists to confuse traffic monitors using DSRC even if the devices or trade them with other people. However, US DOT by requiring DSRC in vehicles may not allow drivers to turn systems off except for specific privacy purposes for which provisions have been made in IEEE 1609.2. Moreover, the Principle of Least Privilege remains a well-respected, fundamental and enforceable information security policy. It dictates that privacy defenders cannot fail to protect location data from one attacker simply because another potential attacker has access. Transportation monitors would work on DSRC systems completely independently from E911 system. Transportation monitors should only be given access to such data if it is deemed important by whoever sets up the transportation system. The same would be true for LBSs such as Waze. If LBS executives were to allow its administrators to casually monitor drivers, the company may be exposed to legal risks. Plus, there may be a moral issue, or even a business perception issue.

This research addresses the following problem: the tradeoff between safety and silence. The safety-silence tradeoff problem is important because society values both safety and privacy. Without a quantitative measure for the tradeoff it is difficult for individuals and policymakers to weigh necessary compromises between the two.

The problem is complicated because VANETs do not work exactly like conventional networks. First, the networks are in motion, following mobility patterns of cars and trucks. Second, the protocols differ. Corporate and home networks, for example, depend on static MAC addresses and TCP/IP, but safety systems in VANETs use new protocols in which MAC addresses are dynamic—they can be changed by the vehicle in transit. In the United Stated the new protocols are part of Dedicated Short Range Communications (DSRC) / Wireless Access in Vehicular Environments (DSRC/WAVE). WAVE Short Message Protocol (WSMP) was created to improve communication speed. Dynamic MAC addressing was added specifically to enable privacy protocols. The BSM for example is part of a protocol called SAE J2735.

The primary contribution of this section is the presentation and proof of the Safety-Silence Tradeoff Equation. Also presented are illustrations and examples of applications of the equation. To our knowledge no method has yet been proposed to compute this tradeoff in vehicular contexts. There are some works on the mathematics of privacy [15] and computational privacy [14], but they do not include discussions of the cost tradeoff between safety and silent periods.

The rest of this chapter is organized as follows. Section 7.1 presents background for the VANET privacy problem, establishing the necessity of silent periods. Section 7.2 presents mathematical analysis and proof. Section 7.3 presents an illustration and Section 7.4 discusses practical considerations. Section 7.5 presents results of simulations to evaluate performance of the equation. Section 7.6 concludes the chapter.

7.1 Silent Periods

VANETs depend on vehicles each having access to accurate Global Positioning System data. Vehicles transmit their positions to each other using vehicle-to-vehicle (V2V) communications. Vehicles may communicate with application servers, such as Location Based Services (LBS) via wired roadside units (RSU) using vehicle-to-infrastructure (V2I) communications. However, as far as their effect on safety is concerned in this chapter, silent periods are relevant only within V2V communications.

Among the earliest and most cited papers [1] in the field of vehicular ad hoc network privacy are recommendations that privacy protection schemes include silent periods. Recent work has suggested privacy protocols be executed at "social spots" such as traffic lights and intersections [16].

These and other recommendations were founded partly on the results of seminal work in mix zones [2]. Silent periods and mix zones operate together. Vehicles cease transmitting and change positions, then begin transmitting again using different identifiers. Silent periods and mix zones have been proven to be effective, and some even suggest in cases where one mix zone is not enough, use multiple cascading mix zones for even more privacy protection [3]. The problem of silent periods and mix zones has been studied for at least the past ten years [6-12].

If the US Department of Transportation mandates the implementation of VANETs, as they are expected to do, not all vehicles will transmit basic safety messages (BSMs). At some point in time, at a given intersection or other potential collision point, perhaps only half of the vehicles will have VANET equipment installed. Even after full

deployment vehicles will go radio silent when implementing privacy protocols. What's the risk?

7.2 Mathematical Analysis

The probability of a crash between vehicles can be estimated by examining the number of possible crash combinations in a given geographical area during a given time frame. Assuming the probability of a crash is different when vehicles are not transmitting BSMs, then given certain assumptions, the probability of at least one crash in a region R during a time Δt is as follows. This is the Safety-Silence Tradeoff Equation. Note: This assumes that the probability is negligible (zero) for a crash between three vehicles. The rationale behind this assumption is that a three-car crash occurs when a third car crashes into a two other cars, not when three cars crash into each other at the same time.

$$prob_{R\Delta t}(crash \geq 1) = 1 - (1 - p_b)^{\binom{b}{2}}(1 - p_u)^{\binom{n}{2}-\binom{b}{2}} \tag{7.1}$$

Proof. Let p_b be the probability of a crash between two vehicles both transmitting BSMs. By the Complement Rule the probability of any particular two such vehicles not crashing is $(1 - p_b)$.

By definition of combination, the total number of potential crashes between any two of n vehicles is $\binom{n}{2}$. Therefore the number of potential crashes between any two of b vehicles is $\binom{b}{2}$, where b is the number of vehicles in R transmitting BSMs during Δt.

By the Multiplication Rule the probability of all b vehicles not crashing is the product of the probabilities of each potential crash not occurring, i.e.

$prob_b(no\ crashes) = (1 - p_b)^{\binom{b}{2}}$.

Define n as the total number of pairs in R during Δt. Then the total number of pairs is $\binom{n}{2}$. The number of possible crashes of unprotected vehicles, vehicle pairs where at least one of the vehicles is not transmitting BSMs, is $\binom{n}{2} - \binom{b}{2}$, and

$$prob_u(no\ crashes) = (1 - p_u)^{\binom{n}{2}-\binom{b}{2}}.$$

This is true because if only one vehicle is transmitting BSMs then the probability of a crash is the same as if neither vehicle were transmitting BSMs. BSMs only work if both vehicles transmit. This analysis assumes that if vehicles have OBUs they transit BSMs, otherwise they do not. With privacy protocols operating on OBUs there are actually four possibilities, (a) both transmitting, (b) both not transmitting, (c) receiving vehicle is not transmitting but other vehicle is transmitting, and (d) receiving vehicle is transmitting but other vehicle is not transmitting. The four-possibility scenario is not considered here however the presentation is easily extendable to more scenarios.

The probability of three vehicles crashing would be the probability of one pair crashing multiplied by the probability of another pair crashing in the same region R during the same time, Δt. Since probabilities, p_b and p_u, are extremely small, it can be assumed that the probability of three cars crashing is negligible. That is, $p_b^2 \approx 0$, $p_b^2 \approx 0$, and $p_b p_u \approx 0$.

The probability of at least one crash, $prob_{R\Delta t}(crash \geq 1)$, is $prob_b(no\ crashes)prob_u(no\ crashes)$, which reduces to (7.1). QED.

7.3 Illustration

Figure 7.1 shows that the maximum number of possible collision combinations rises exponentially with the number of vehicles. If we choose n=10 we can vary b to see

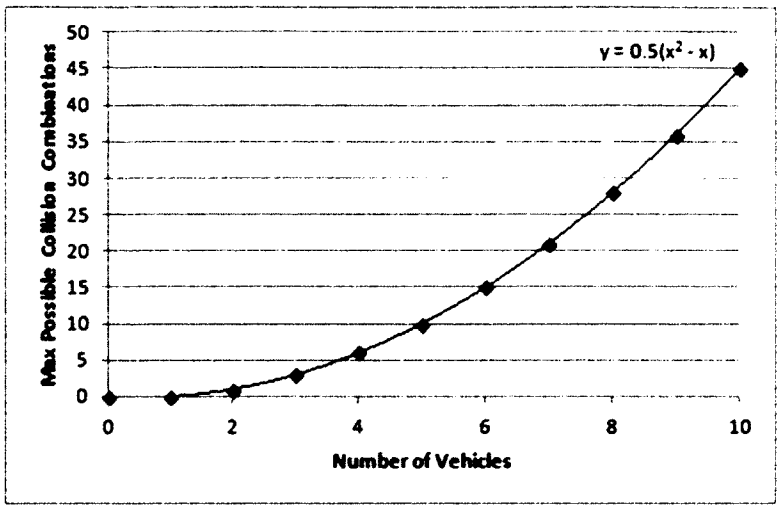


Figure 7.1: Maximum possible collision combinations. Maximum rises exponentially with the number of vehicles present at the intersection.

the relationship between *b*, the number of vehicles transmitting BSMs, and *n*, the total number of vehicles in region *R*.

Figure 7.2 shows that if 5 of 10 (50%) vehicles transmit BSMs, 35 out of 45 (78%) collisions remain possible. One could conclude that if a large percentage of vehicles would enter a silent period, then neighboring vehicles might do likewise since the marginal loss in collision protection would be at a minimum.

7.4 Practical Considerations

With four vehicles at a four-way stop, there are 28 possible impact points but $\binom{4}{2}$ =6 possible collisions between two vehicles. See Figure 7.1 The equation does not count the number of potential impact points, rather, it counts only the number of combinations of pairs of vehicles.

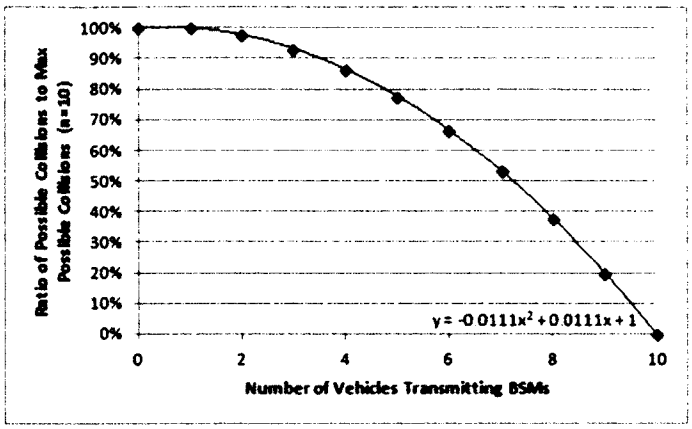


Figure 7.2: Relationship between BSMs and potential collisions. An inverse exponential relationship exists between the proportion of vehicles transmitting BSMs and the proportion of potential collisions unprotected by BSMs. Under assumptions, if 5 of 10 (50%) vehicles transmit BSMs, 35 out of 45 (78%) collisions remain possible.

The safety-silence tradeoff equation predicts the probability of at least one collision is $1 - (1 - p_b)(1 - p_u)^5$ where p_b and p_u are as defined in Section 7.2 and assume n=4 and b=2.

With six vehicles, an intersection of one two-lane road and one four-lane road, there are 56 possible impact points assuming expected traffic flow. Three possibilities are ruled out in the model: A cannot collide with E, C cannot collide with F, and E cannot collide with F. Therefore, there are $\binom{8}{2} - 3 = 12$ possible collisions between two vehicles.

Suppose A, B and C do not transmit BSMs, but D, E, and F do. Then

$prob_b(no\ crashes) = (1 - p_b)^{\binom{3}{2}-1} = (1 - p_b)^2$ and

$prob_u(no\ crashes) = (1 - p_u)^{\binom{8}{2}-3-\binom{3}{2}} = (1 - p_u)^{10}$.

In sum, impossible collision combinations must be subtracted from the exponent in the safety-silence tradeoff equation. This can be accomplished by including constants. Let

the quantities, c_b and c_n, be the number of logistically impossible collision combinations between vehicles both transmitting BSMs and between all vehicles, respectively. Then the modified equation would be as follows.

$$prob_{R\Delta t}(crash \geq 1) = 1 - (1 - p_b)^{\binom{b}{2} - c_b}(1 - p_u)^{\binom{n}{2} - c_n - \binom{b}{2} + c_b} \quad (2)$$

7.5 Simulations

Results were compared for two types of intersections, Class 1 and Class 2, as pictured in Figure 7.3 and Figure 7.4, respectively. All possible values for *b* and *n* were computed and simulated. No combinations were removed using constants as in (2); all computations were pure, as in (1).

Probabilities and other simulation parameters were set according to Table 7.1. There were 1000 simulation runs per scenario.

Comparison of computed versus simulated results showed that the equation estimated simulation results accurately. However, as probabilities become increasingly small computations become less accurate. Numerical methods will need to be employed in order to maintain numerical precision.

Figure 7.5 shows the results for Class 1 intersection simulations with probability scenarios *A*, *B* and *C* as defined in Table 7.1. Figure 7.6 shows results from Class 2 intersection simulations. Figure 7.5 and Figure 7.6 echo a pattern similar to that shown in Figure 7.1, except the reduction in the probability of at least one crash occurs even more gradually with the increase in the ratio of the number of vehicles transmitting BSMs to the total number of vehicles in the system.

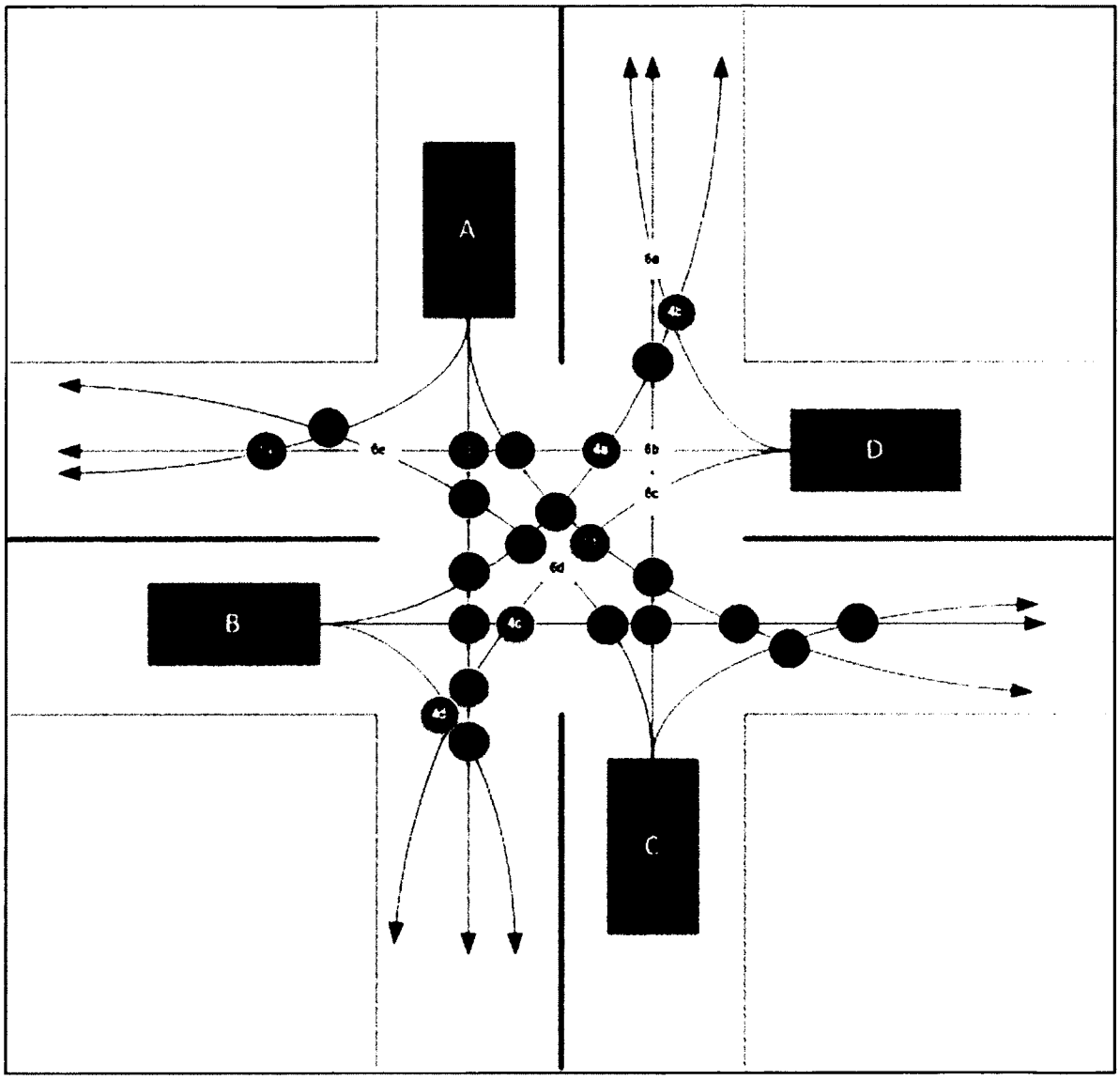


Figure 7.3 Four vehicles at an intersection of two two-lane roads

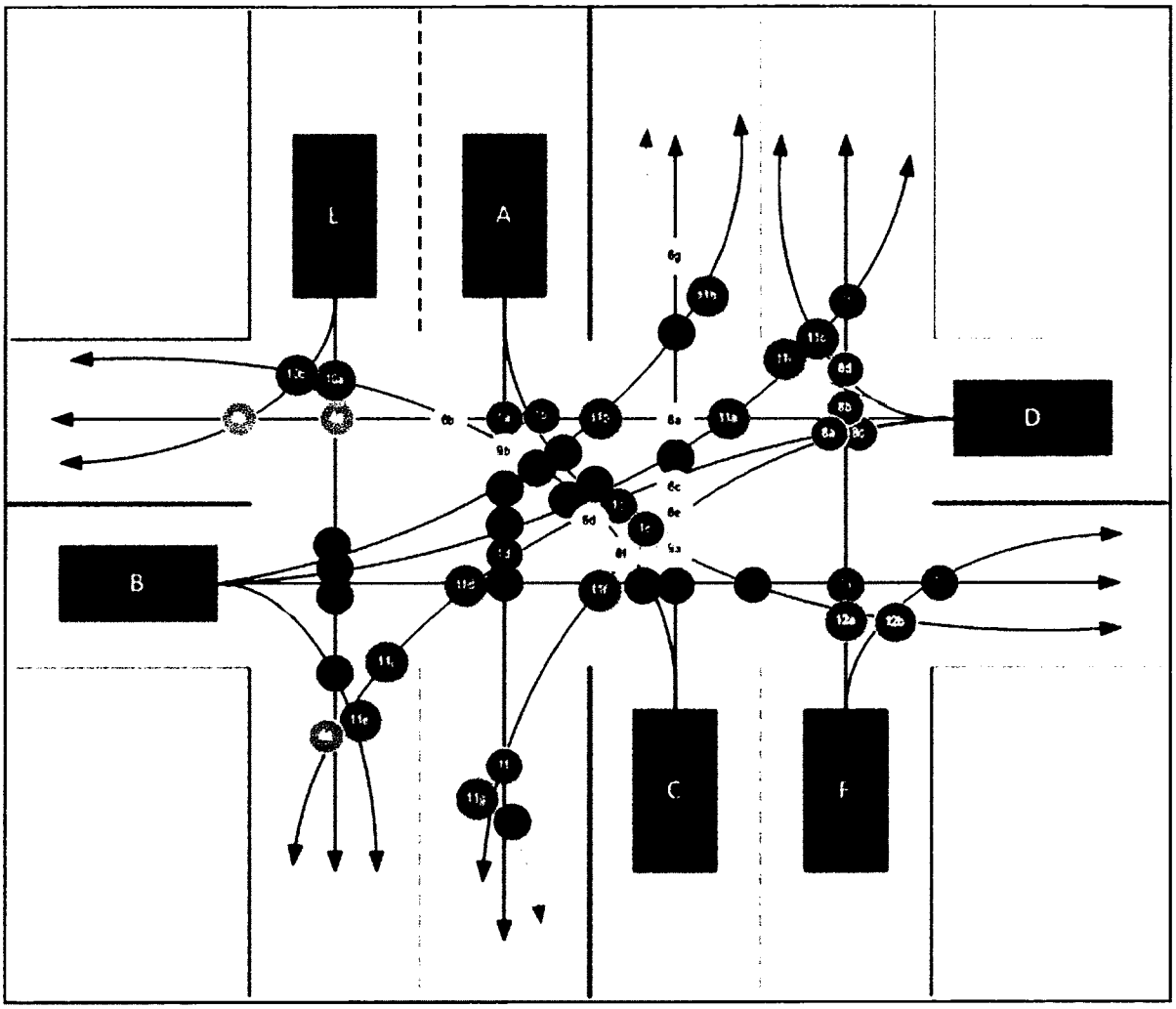


Figure 7.4: Six vehicles at an intersection of a two-lane and a four-lane road. Class 2 simulation scenario.

Table 7.1

Safety-Silence Tradeoff Simulation Parameters

Parameters	**Values**
Intersections, vehicle quantities (n) and vehicles sending BSMs (b)	Class 1: n=4 vehicles at an intersection of two two-lane roads, $0 \leq b \leq 4$ Class 2: n=6 vehicles at an intersection of a two-lane and a four-lane road, $0 \leq b \leq 6$
p_b, p_u	Scenario A: $p_b=.1$, $p_u=.2$ Scenario B: $p_b=.01$, $p_u=.02$ Scenario C: $p_b=.001$, $p_u=.002$

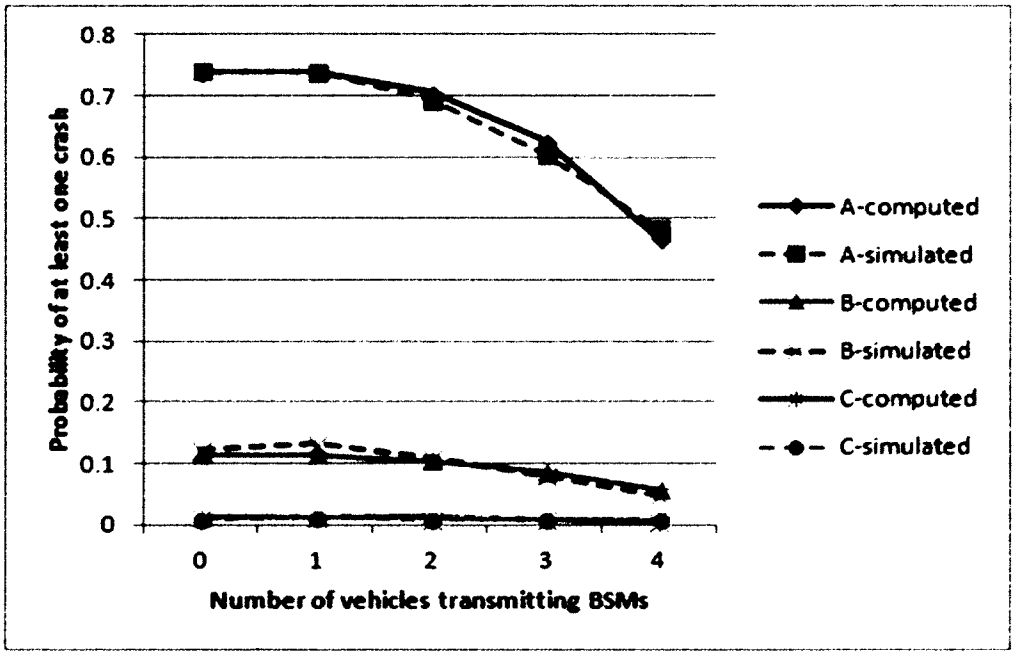


Figure 7.5 Intersection Class 1, probability scenarios A, B and C

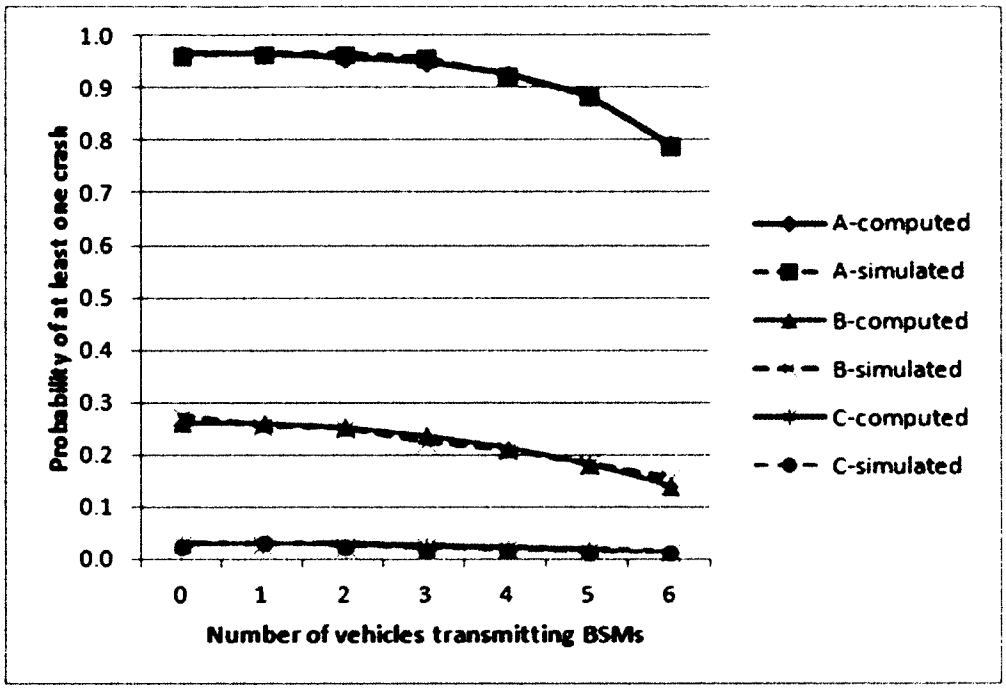


Figure 7.6: Intersection Class 2, probability scenarios A, B and C

The figures assume intersections are full, i.e. all possible positions of vehicle in an intersection have a vehicle present and waiting to enter the intersection. It may be more likely that one or more possible positions are unfilled. Consider the diagram in Figure 7.3. It may be that positions *A*, *B* and *C* are occupied, but *D* is vacant; there is no car at position *D*. In this case all vehicle combinations involving D cannot produce crashes. A simulation was run using a Class 1 intersection under Scenario *A* as defined in Table 7.1.

Simulations were run using extremely high probabilities because simulations using probabilities in the realistic range would not hold numerical precision. However, computed and simulated results showed clearly similar patterns. If these patterns hold in smaller probabilities then the Safety-Silence Tradeoff Equation may serve as a useful back-of-the-napkin estimator. For a perpective using a broad set of vehicle densities, see Figure 7.7.

About 40% of all accidents in the US occurred at intersections in 2006, 8,500 of which were fatal and 900,000 of which were injurious [17]. In Japan, in 1997, 60% of accidents occur at intersections. Based on a study of 150 four-legged intersections in the Tokyo metropolitan area, researchers observed that the average probability of a vehicle encountering an obstacle vehicle was 0.339 and the average probability of a following vehicle driver's failure was 0.0000002 [18]. These vehicles were not outfitted with on-board units (OBUs), the devices that transmit BSMs. No data are available regarding accident probabilities of vehicles with OBUs installed, however one recent NHTSA report estimated that V2V could one day address 79% of vehicle crashes [19]. From this

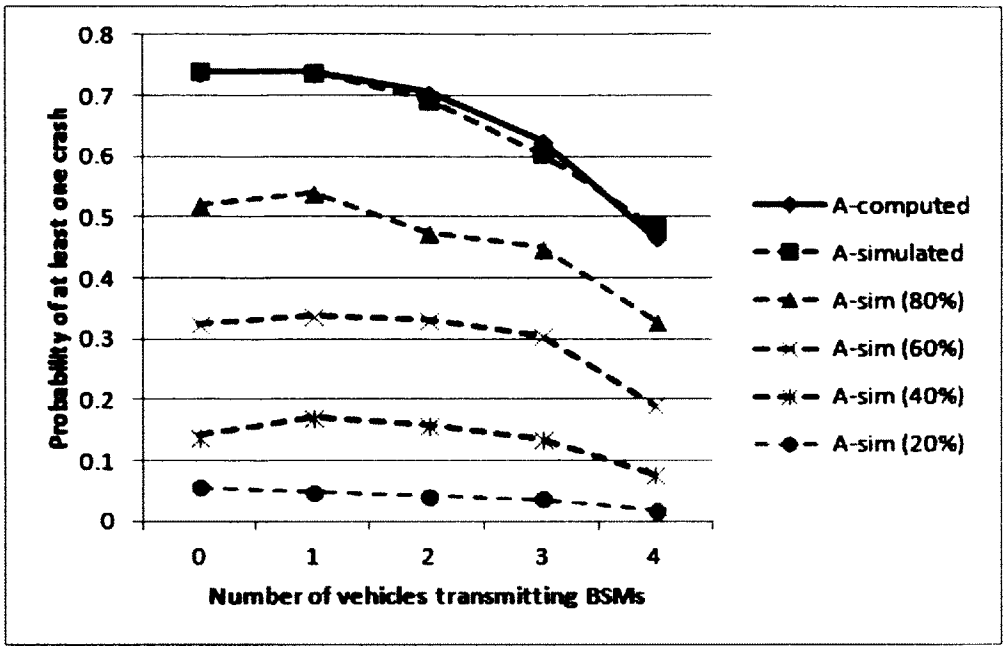


Figure 7.7: Class 1 intersection at five levels of intersection density. All run with with simulation Scenario A: 100%, 80%, 60%, 40% and 20%. The results show a near-linear relationship between the percentage of vehicles positions occupied at a 4-way stop intersection and the probability of at least one crash at that intersection.

we can roughly estimate that the probability of a crash at an intersection without any BSMs is 0.339 times 0.0000002, or 0.0000000678, and the probability of a crash at an intersection with all vehicles transmitting BSMs is (1-0.79) times 0.0000000678, or 0.00000001423.

7.6 Summary

This chapter has shown the following. First, an inverse exponential relationship exists between the proportion of vehicles transmitting BSMs and the proportion of potential collisions unprotected by BSMs. This is a straightforward computation of combinations of crashes possible between vehicles in a region. The conclusion from this

observation is, if privacy protocols must include a silent period, then perhaps as many vehicles as possible should execute privacy protocols when any vehicles decide to go transmission silent. This is because the marginal loss in collision protection is greatest when one single vehicle decides not to transmit BSMs while all others do.

Second, the Safety-Silence Tradeoff Equation can estimate the cost of VANET privacy methods that require silent periods. This can be done by computing the likelihood of crashes between combinations of vehicles at intersections. In order to accomplish this in practice, future work would necessarily include gathering statistics to obtain accurate probabilities in different intersection situations and developing numerical techniques to maintain computational precision when applying the probabilities, p_b and p_u. The specific probabilities, p_b and p_u, will likely differ given a range of factors including location, weather conditions, and the types of vehicles in the region.

Third, in practice some combinations of potential collisions might be eliminated. The equation can be modified to accommodate these scenarios by the inclusion of certain constants. Careful analysis of historical vehicle crash data would be required in order to accomplish this in practice. There may even be a need to develop a handbook or other database for crash probabilities between vehicles at different types of intersections, including which combinations can be ignored.

Fourth, traffic density affects the computation. If intersections are not full of cars, probabilities of crashes are lower. Sophisticated traffic management systems of the future, armed with hyper-accurate historical data of past crashes, and vehicle densities, might be able to inform vehicles seeking privacy using silent periods of the optimal (safest) times and locations to execute the protocols.

Finally, transmission silence could occur because the dissemination of on-board units (OBUs) has not reached a high level of saturation in a particular geographical area. During this ramp-up interval the use of privacy protocols which require silent periods would be less risky than in intervals or locations of full saturation.

CHAPTER EIGHT

EVALUATING PROTOCOLS USING KDT

Privacy protocols have to date generally measured privacy performance based on metrics such as *k*-anonymity. Because of the frequency and precision of location of queries in vehicular applications, privacy measurement may be improved by considering additional factors. As it is described in Chapter Three, this applies KDT anonymity, a composite metric including average anonymity set size, K, average distance deviation, D, and anonymity duration, T. This section evaluates five privacy protocols under realistic vehicle mobility patterns using KDT anonymity, and compares KDT anonymity with prior metrics.

The simulation set-up is summarized in Table 8.1. In each simulation, the goal was to achieve anonymous FPL LBS access, maximizing KDT. That is, vehicles attempted to anonymize with as many other vehicles as possible, remaining anonymous for as great a distance as possible, and for as long a time as possible. Each vehicle did the following.

1. Enter region, deanonymized.
2. Execute privacy protocol, become anonymized.
3. Exit region and become deanonymized again, or otherwise terminate anonymous LBS access.

The rationale behind the anonymization process outlined above is to create a controlled environment to compare protocols. One could imagine entering a large region where there are only a few means of ingress and egress, with each ingress/egress point

Table 8.1

KDT Simulation Parameters

Parameter	Setting
Size of region, R	3000 m x 3000 m
Communications Range	200m, 300 m, or 400m
Mobility Models	GMSF (City, Urban, Rural) [23]
Mix Points	50, 60, 70, ..., 150, or continuous
Silent Period (Δt)	30 seconds
Simulation Time	2000 s (33.3 min)
Avg. Vehicle Speed	20 m/s

monitored by LPRs. If a motorist desired to move about the interior of region anonymously she would have to anonymize after the point of ingress and would lose anonymity upon egress.

8.1 Threat Model

Intelligent Transportation Systems (ITSs) are expected to one day include air-traffic-control-like LBS systems called Traffic Management Systems (TMSs) which would enable traffic managers to guide drivers or driverless cars based on current vehicle traffic conditions. If US DOT decides to mandate the use of TMSs, and if vehicles access TMSs through WSMP, then a malicious insider or an outsider who has hacked a TMS might be able to track vehicles from the comfort and anonymity of her own home laptop.

This chapter analyzes privacy protocols using the threat modeling framework proposed in [7], which defines the means, actions and goals of the potential attacker. This

paper assumes a global passive adversary (GPA) with the following means: access to LBS application data, RSU data and perhaps to certain license plate reader (LPR) or other camera data; and knowledge of geography (road maps / road topology), traffic conditions (blocked / slow roads), home owner names and addresses and geographical coordinates, and the target's name, address, license plate number, and perhaps expected mobility profile. This paper considers DSRC communications only, not cell phone data or other information from other devices even though that might also be useful to an attacker.

The GPA's actions are assumed to be passive; that is, the GPA would eavesdrop only and would not alter the data being transmitted. The scope of the attack would be global; that is, the GPA could observe data over a wide region, the entire area covered by the TMS. The temporal scope of the actions may be long term; that is, the GPA could eavesdrop for hours, days, months, or even longer periods of time.

The goal of the GPA would be to determine whether a specific vehicle (target) was at a given place at a given time in order to track the target in real time.

8.2 Mobility Patterns

This research employed Generic Mobility Simulation Framework, GMSF [21], which offers Multi-agent Microscopic Traffic Simulator, MMTS, trace files linked on the GMSF website [22] and provided at the Laboratory for Software Technology website [23], specifically city, urban and rural. All three trace files contain records of time-stamps, vehicle-ids, x-coordinates, y-coordinates within a 3000x3000 meter (9 million square meter) grid. Each mobility pattern starts with a different number of vehicles, v. City starts with v=897. Urban starts with v=488. Rural starts with v=110.

Vehicles enter and leave the region at the same rate, but the number of vehicles in the pattern at any given time is not always precisely the same as the number at the start.

Sometimes road topologies, such as in the Freeway pattern (a straight road with perhaps several lanes) and the Manhattan pattern (a grid of horizontal and vertical roads), provide wide ranging linear density versus area density. That is, the vehicle density per linear meter can be out of sync with the vehicle density per square meter, especially when compared with more realistic road topologies. For example, for 900 vehicles in a 3000 m x 3000 m grid, the Freeway pattern might have a linear density of 0.3 v/m, 900 vehicles divided by 3000 meters, and a square density of 0.0001 v/m^2, 900 vehicles divided by 9 million square meters. The Manhattan pattern would have a linear density of 0.004839 v/m, 900 vehicles divided by 186,000 meters, but the same square density as the Freeway pattern. In other words, the linear density of the Manhattan pattern is 1.6% that of Freeway pattern given the same square density.

The mobility patterns used in this simulation, however, have similar linear distances: city, 14,783 meters; urban, 13,955 meters; and rural, 10,175 meters. The areas covered are identical, 3000 m x 3000 m, so the mobility patterns provide relatively realistic traffic flows and comparable roadway linear distances and square areas.

8.3 Location Privacy Protocols and Mix Points

To create mix zones, as defined in [24], simulations used the concept of a *mix point*, a position in space and time, with coordinates *(x,y,t)*. The mix point was used to create a circular mix zone of radius, *r*. If a vehicle was positioned within *r*, i.e. within *comrange*, of *(x,y)* at time *t*, then that vehicle initiated whatever the privacy protocol required. If the vehicle never came within comrange of any mix point then the vehicle

never became anonymized. This simulation tested five protocols: SMP-R, stationary mix points, occurring at regular time intervals; SMP-I, stationary mix points, occurring at irregular time intervals; OTFP-R, randomly chosen on-the-fly mix points, occurring at regular time intervals; OTFP-I, randomly chosen mix points, occurring at irregular time intervals; and GLRP-2, group leader relay points, which occur continuously throughout the trajectory of a vehicle designated as the leader of a group of vehicles traveling within comrange of the leader. The number, 2, in GLRP-2 indicates that vehicles join the group in pairs. Performance was evaluated against the theoretical values computed in Section 3. This section provides descriptions of each protocol.

8.3.1 Stationary Mix Point (SMP)

An SMP creates a region that does not move in which vehicles may switch pseudoIDs. A similar protocol is described in [55], but in the SMP model presented here, a fixed point *(x, y)* was chosen. To maximize k, the busiest intersection in the mobility model was chosen as the "social spot" for mixing. In scenario SMP-R, regular time intervals were chosen. In scenario SMP-I, irregular time intervals were chosen. Vehicles that were within radius, r, of point *(x, y)* at time, t, were added to the anonymity set.

Upon anonymizing, vehicles enter a silent period. They cease all communications at both MAC and APP layers because, if they were to continue communications via one, under RSU LBS collusion they would be linkable to the other. All vehicles in the anonymity set changed pseudoIDs, but remained silent until the silent period expired, at which point all silent vehicles resumed communications, including anonymous LBS access, using new identifiers. See Figures 8.1 and 8.2.

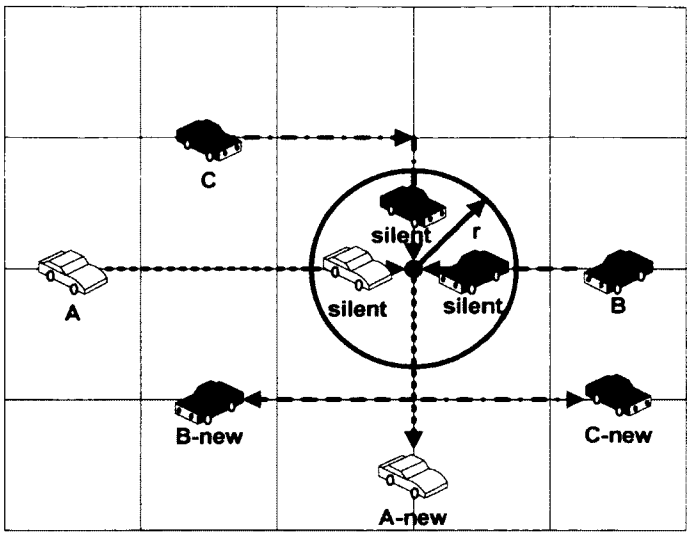


Figure 8.1: Stationary mix point (SMP) VANET privacy protocol. The circular mix zone does not move. At regular, predetermined points in time, a beacon announces it is time to mix. Vehicles in comrange cease transmissions, switch pseudoIDs, and resume communications after a silent period.

Time	Beacon	Vehicle	Attacker
0		Enters region	Marks, tracks vehicle
1	Announces SMZ		
2		Enters SMZ	
3		Goes silent	Loses vehicle
4		Switches pseudoIDs	
5		Resumes FPL LBS	Sees anon. set of vehicle
6		Reaches boundary	Re-marks vehicle

Figure 8.2: Roles of entities in the SMP models (SMP-I and SMP-R)

Simuations varied vehicle density (vehicles per meter of road, or vehicles per square meter of area, as exhibited in mobility models), mobility model (city, urban, rural), the number of mix points (from 50 to 150), and comrange (r=200, 300 and 400 meters). Silent period duration was a constant 30 seconds. There were 2000 seconds in each simulation run, so if there were 50 mix points, then there were 40 seconds between each mix point, if time intervals were regular. Again, mix points are points in space and time, with coordinates *(x,y,t)*, so for SMP-R, a series of mix points would have constant values of x and y, varying only t.

The key strength of the SMP model is that locations can be strategically chosen to achieve the best privacy. If especially opportune social spots are selected as in [55] then SMP will be more effective than other methods which may attempt to anonymize vehicles at less opportune locations.

The SMP model depends on a trustworthy anonymization location. However, if the zone were widely known, attackers could simply place LPRs near the SMP and completely defeat privacy. Social spots suggested in [55] might be vulnerable since eavesdroppers might deploy LPRs at the social spots.

If LBS requires authentication, and there is RSU collusion, then FPL LBS anonymity could be protected using blind signatures. Blind signature schemes are outside the scope of this paper. For more information refer to [49] and [59].

8.3.2 Group Leader Relay Point (GLRP)

The group leader model has been presented in several important papers, notably [7] and [52]. In this paradigm one vehicle is the designated coordinator, or group leader (GL), of a cluster of vehicles which travel together. In the AMOEBA model [52] when

two vehicles come within range of a GL they go silent as in the SMP model. But the mix zone is not stationary. It moves with the GL, centered on the GL. Vehicles must stay within comrange of the GL to communicate anonymously with the LBS. See Figs. 30 and 31.

The variables for the GLRP-2 model are the same as for the SMP model except anonymity set size for any anonymized vehicle is 2 because vehicles anonymize in pairs. Also there is no fixed point in space or time for beaconing join/mix requests. The group leader beacons frequently along its trajectory.

The key strength of the GLRP model is that the GL acts as a TTP to protect vehicles requiring anonymity. GLRP also covers a wider geographical area than the SMP model because mix points traverse a region, the path of the GL. See Figure 8.3.

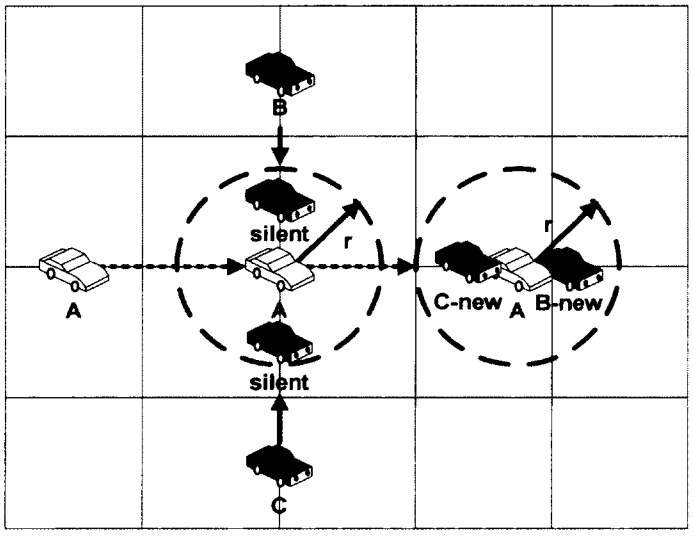


Figure 8.3: Group leader relay point (GLRP-2) VANET privacy protocol. The circular mix zone moves, centered on the group leader, labeled A. The group leader is never anonymized and never goes silent, but continuously beacons for other vehicles to mix in pairs. The group leader relays mixed vehicles' LBS requests as long as they are in communications range.

The GLRP model depends on two things: trustworthy GL volunteers and other vehicles with which to anonymize who will travel within communications range of the GL. The roles of the entities are shown in Figure 8.4. In sparsely populated areas it is unclear whether there would be a sufficiently large number of vehicles both willing and able to provide meaningful privacy under these conditions. Some researchers suggest the group model is infeasible due to limitations of bandwidth and computation power, since pseudoID schemes create large certificate revocation lists, CRLs, and associated checking costs, network overhead necessary to verify that certificates have not been revoked [9].

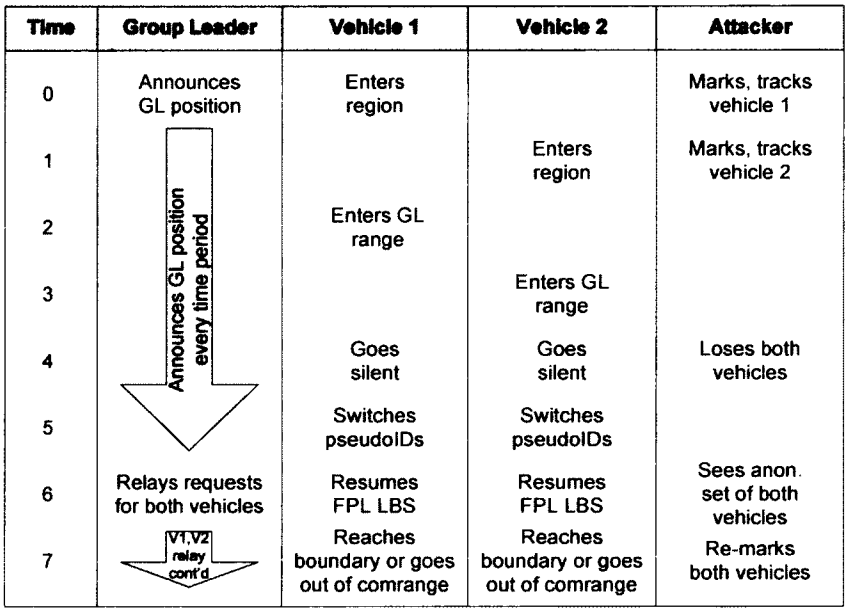


Figure 8.4: Roles of entities in GLRP-2 model. Group leader continuously beacons for vehicles to mix in pairs.

Joining the group requires some communications overhead, too. The establishment of a group requires communications that do not provide any direct service. To illustrate, suppose there is a GL transmitting with pseudomym A. Then a vehicle enters the group network, transmitting with pseudonym B. Another vehicle enters the network, transmitting with pseudonym C. If B and C go silent, change identifiers, then begin transmitting after a time, Δt, then they would be anonymized. AMOEBA uses the random silent period for joining the group, as presented in [53]. If vehicles join groups in pairs, then this ensures an anonymity set size of 2 for each anonymized vehicle.

Communicating to LBS while a member of a group requires a great deal of overhead. Under the AMOEBA protocol, GLs relay LBS communications for the vehicles in their group network. This effectively doubles every vehicle's LBS overhead. It is possible that one GL may handle many pairs of of anonymized vehicles. This could create a communications bottleneck point at the GL.

LBS authentication also requires additional encryption in the group relay model. For example, in the AOSA system [58], a vehicle, V_i, sends a signed service request, m, to LBS, via a group leader, GL, a roadside unit, RSU, and a proxy server. The steps of the AOSA process are as follows.

1. V_i sends to GL the signed service request, m, represented by $V^-_i(m)$ and his digital certificate, $Cert(V_i)$ encrypted by the LBS public key, LBS^+ and by a previously established secret key, GL_i^S. This can be represented as $GL_i^S\left[LB^{S+}(V^-_i(m), Cert(V_i)\right]$, or more simply, as $GL_i^S[M]$.
2. GL decrypts M and adds the GL's location, LocGL, and GL's digital certificate, $Cert(GL)$, then signs the message with the GL's private key, GL^-,

resulting in $MGL = GL^{-}(M), Cert(GL), LocGL$. GL then forwards to the roadside unit, RSU.

3. RSU relays message to Proxy.
4. Proxy verifies GL's certificate and forwards request to *LBS*.
5. The *LBS* (a) decrypts the message with its private key, (b) verifies the vehicle's identity and (c) verifies the vehicle's authority for the service requested. If all is satisfactory, service provider sends a reply, *R*, which includes session key, LBS_i^S for communication between the vehicle and *LBS*. This results in a reply message encrypted with the vehicle's public key, then with GL's public key, or $GL^{+}(V^{+}_i(R))$.
6. Proxy verifies LBS's certificate and forwards reply to RSU.
7. RSU relays reply to GL.
8. Finally, the GL forwards reply to vehicle using group secret key. The final reply may be represented as $GL_i^S[V^{+}_i(R)]$.

If a system administrator knew the private key of the LBS, AOSA protocol would not protect anonymous LBS access for a vehicle because LBS receives the digital certificate of V_i and so would know the identity of V_i. Another method, such as a blind signature scheme, may be required to protect anonymity if LBS requires authentication.

8.3.3 On-the-fly Point (OTFP)

The OTFP model, similar to that presented in [50], Privacy-by-Decoy (PBD), is similar to SMP except vehicles choose anonymization points at random locations. With respect to choosing the anonymyzation location, the choice of timing could be at regular intervals or irregular ones. If regular time intervals were instituted then there would need

to be some method of informing the vehicle as to what the timing would be. If irregular time intervals were instituted, the vehicle could beacon for anonymity at any opportune point along its trajectory. See Figures 8.5 and 8.6.

As in the other models, in the OTFP model, when vehicles willing to anonymize move within communications range of each other, they agree to anonymize, go silent for a time, then resume transmissions.

The circular mix zone moves with the vehicle desiring privacy, a.k.a the pirate. Pirates beacon for vehicles willing to mimic their FPL LBS requests, a.k.a. parrots. When a pirate finds one or more parrots, all vehicles cease transmissions, switch IDs, and resume after a silent period. The pirate resumes FPL LBS requests normally, but the parrot uses the pirate's encrypted credentials to access the LBS on behalf of the pirate.

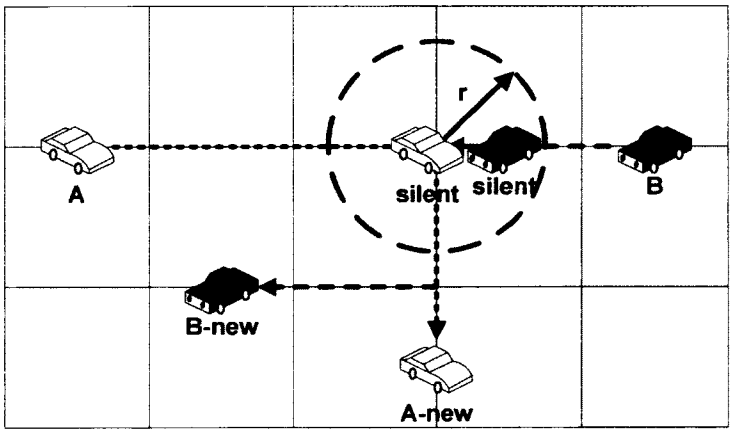


Figure 8.5: On-the-fly point (OTFP) VANET privacy protocol. The circular mix zone moves with the vehicle desiring privacy, a.k.a the Pirate. Pirates beacon for vehicles willing to mimic FPL LBS requests, a.k.a. Parrots. When a Pirate finds one or more Parrots, both vehicles cease transmissions, switch pseudoIDs, and resume communications after a silent period. The Pirate resumes FPL LBS requests normally, but the Parrot uses the Pirate's encrypted credentials to access the LBS on behalf of the Pirate.

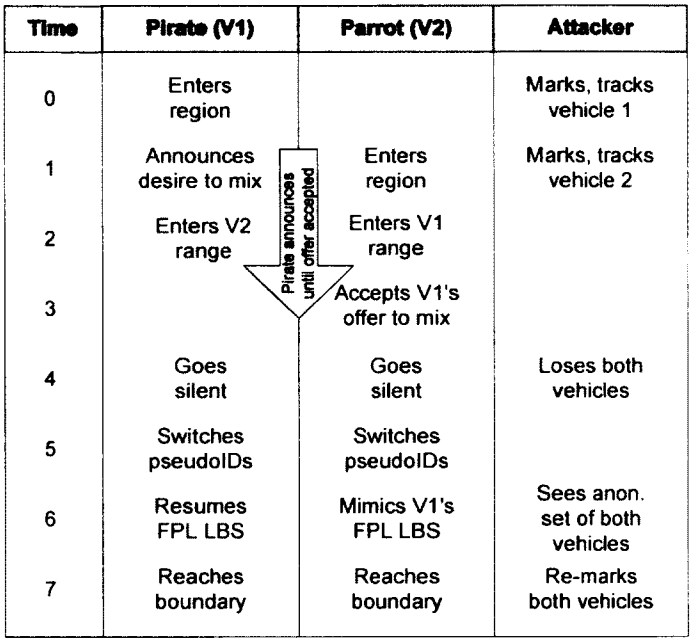


Figure 8.6: Roles of entities in OTFP models. Pirate continuously beacons until one or more Pirates are found.

The variable parameters for the OTFP model are identical to the SMP model except the location of mix points is not fixed. Rather, mix zone locations are determined randomly, on-the-fly, as potential Parrots and unanonymized Pirates move within comrange of each other.

Randomly located mix zones makes it more difficult for attackers effectively to set up deanonymizing equipment like LPRs since the locations of the mixes are not fixed. The OTFP model offers more user control, since individual vehicles rather than a centralized authority would initiate mix requests.

The OTFP model requires overhead to establish the Pirate-Parrot relationship. It also increases the number of LBS requests by a factor equal to the number of Parrots. In

the SMP and GLRP models, FPL LBS users connect anonymously. The in those models the assumption is the LBS cannot distinguish one request from another. If a TMS permits only one type of request, then the nature of the query will not distinguish the vehicle from its neighbors. The OTFP model offers a novel method for vehicles to remain anonymous even when LBS requires authentication even by identifiers such as userID. Parrots protect Pirates by accessing LBS using Pirates' credentials and the Parrots' locations.

Normally, any vehicle, V_i, would send a request to an LBS by encrypting a signed message, m, with the LBS public key, such as $LBS^{+}(V^{-}_i(m), Cert(V_i))$. The message may contain both the request and authentication credentials, for example, username, U, password, P, query, Q, and, most importantly, location, Loc. That is, $m=U,P,Q,LocV_i$.

Under OTFP, as described in PBD, vehicles separate authentication credentials and query from location. LBS would be constructed in such a way as to allow this. So m would be divided into $m1=U,P,Q$ and $m2=LocV$, where $LocV$ could be the location of any vehicle, but is specifically designed to be the location of either a Pirate, V_i, or a Parrot, Vj, that is to say, $LocV_i$ or $LocV_j$.

Both Parrots and Pirates would communicate with LBS with none of the additional group management overhead required by the GLRP model. There is additional overhead, though, because messages sent to and from the LBS and the Parrot would not be readable by the Parrot and would be discarded. (Only the Pirate could read the LBS replies, since they would be encrypted with the Pirate's public key.) The additional overhead with parroting would linearly increase the amount of network traffic.

8.4 Performance Evaluation Using Desired Properties

This section presents the relative performance of the location privacy models using the kdt metric and twelve common VANET privacy protocol desired properties, listed in Table 8.2 and described in Chapter Two. Property 8 in Table 8.2 is evaluated using anonymity-related metrics, as described by (3.1), (3.2) and (3.3) as well as our proposed metric, KDT, described in Chapter Three.

Since the models anonymized roughly the same number of vehicles (see Figure 6.2) and the silent period was constant the models performed equivalently with respect to their impact on safety.

Privacy protocols diminish authentication effectiveness when they introduce vulnerabilities or impair performance in processing or accessing digital signatures. Let us define the latter as public key infrastructure (PKI) efficiency. In each model except GLRP-2 it was assumed that there would be multiple certificates per on-board unit (OBU) as in [63]. In GLRP-2 it was assumed that each vehicle belonging to a group would have that same number of certificates as the other models, plus additional group signature. The V2V communications necessary to establish the group, plus the communications necessary to relay and convert messages to the LBS, represent additional impairment of PKI efficiency and therefore the GLRP-2 model was rated lower than the other models.

Privacy protocols diminish pseudonymity when they risk linkage between pseudoID and VIN or PII. There is a natural tradeoff between authentication by digital signature and pseudonymity by pseudoID. For authentication to be fast and efficient, the fewer the pseudoIDs and certificates the better; for pseudonymity, the more the better.

Table 8.2

Desired Properties and Location Privacy Models

No	Property	G-2	S-R	S-I	O-R	O-I
1	Collision Avoidance	≈	≈	≈	≈	≈
2	Authentication	○	≈	≈	≈	≈
3	Pseudonymity	●	≈	≈	≈	≈
4	Untrackability	≈	≈	≈	≈	≈
5	Untraceability	●	≈	≈	≈	≈
6	Accountability	○	≈	≈	≈	≈
7	Revocability	○	≈	≈	≈	≈
8	Anonymity	*	*	*	*	*
9	Decentralization	●	≈	≈	≈	≈
10	Undeanonymizability	○	○	○	●	●
11	Efficiency	○	●	●	○	○
12	User Control	≈	○	○	●	●

● indicates superior performance compared to other privacy models

○ indicates inferior performance compared to other privacy models

≈ indicates equivalent performance compared to other privacy models

Abbreviations: G-2 (GLRP-2), S-R (SMP-R), S-I (SMP-I), O-R (OTFP-R), O-I (OTFP-I)

Consequently, for the same reason that the GLRP-2 model was rated lower than the other models in terms of Property 2, it was rated higher than the other models in terms of Property 3. That is, additional pseudonyms slow down the authentication process, but they at the same time add an additional layer of pseudonymity.

Untrackability is accomplished, not by pseudo-identity, but by mixing. (Refer to the discussion on unlinkability in the prior subsection.) In sum, all models are equivalently untrackable

To the extent that group signatures add an additional layer of identity privacy via pseudonymity, group signatures provide better protection against tracing.

All of the simulated protocols require identities and pseudo-identities with digital certificates issued by a CA which could deanonymize the transmission if, for example, required to do so by a court order. The group model provides for both a CA to enable single-entity digital signatures and a group manager to enable group signatures. All models provide the same technical level of accountability. It is assumed that it would be more difficult for a law enforcement authority to obtain a warrant and receive deanonymized vehicle information from two entities than from just one, so the group model is rated inferior to the others because it includes both the CA and the group leader to cooperate in order to achieve accountability.

Some researchers suggest the group model is infeasible due to limitations of bandwidth and computation power, since pseudoID schemes create large certificate revocation lists, CRLs, and associated checking costs, network overhead necessary to verify that certificates have not been revoked [9].

Privacy protocols diminish anonymity when they do not provide high levels of indistinguishability between pseudoIDs, as measured by (3.1), (3.2) and (3.3). Equations (3.7) and (3.12) measure distance deviation and anonymity time, respectively. Equations (3.1), (3.7) and (3.12) are used to compute the composite metric, *kdt*. Table 8.3 provides details and simulation results for this property.

Privacy protocols diminish decentralization when they fail to call for multiple independent TTPs. Any protocol could implement blind signatures but the group model explicitly requires at least two separate credential-issuing entities (TTPs) so GLRP-2 is evaluated higher than its peers in this regard.

Combination MAC-layer/APP-layer defenses, such as endpoint protection zones (EPZ) [28] could enable undeanonymizability even when LBS and RSU collude. Such defenses could be used in combination with any protocol in this study, however EPZ defends endpoints only; it does not not defend against LPRs at region border ingress/egress points.

Dummy event methods [11][12][13][14][15] have been researched but none, except the OTFP-class model, PBD [50], has explicitly addressed the RSU-LBS collusion problem beyond the basic defense offered by EPZ. PBD not only protects against RSU-LBS collusion, but also enables LBS login. This is the strongest undeanonymizability defense of which we are aware, especially considering it requires no anonymization server. The other models offer no protection against deanonymization from cross-referencing or collusion.

In the case of the group model, announcements are frequent; they occur every time period. Let T represent the total number of time periods that a GL is in a region. In

the case of the SMP and OTFP models, the optimal duration is $^{T}/_{S}$, where S is the duration of the silent period.

Group and on-the-fly models require vehicles to respond to announcements with a join request. As shown in Figure 6.2, the number of vehicles anonymized is roughly the same between models. Let N be the number of anonymized vehicles and let us assume one additional message for each join request. The group model must also generate keys for each of the N vehicles anonymized.

Group model requires GL to respond to join request with a join confirmation. In the OTFP model the pirate provides the join confirmation message. So both GLRP-2 and OLTP models require N more messages.

GLRP-2 requires GL to relay every LBS request, effectively doubling the number of LBS-related transmissions. Assume each of the N anonymized vehicles are anonymized for half the time the GL is in the region. Further assume FPL LBS requests occur every time period. In that case the number of additional LBS relays would be *NT/2*.

Finally, OTFP requires parrots to send false LBS requests, decoy messages, every time period. Under the same assumptions as the prior paragraph, the dummy requests would be *NT/2*. Table 8.3 shows that the SMP models are more efficient.

SMP models provide little flexibility for privacy protocol users, as anonymization points and times are set by central authorities, not motorists. GLRP provides some motorist-level control as the group leader is itself both an anonymization point and a motorist. However, GLs cannot anonymize, and vehicles desiring anonymity have little control over where and when any GL will be available. In fact GLs could themselves be

Table 8.3

Network Overhead and Location Privacy Models

No	Communication	G-2	S-R	S-I	O-R	O-I
1	*Announcement*	*T*	*T/S*	*T/S*	*T/S*	*T/S*
2	*Join request*	*N*	*0*	*0*	*N*	*N*
3	*Key generation*	*N*	*0*	*0*	*0*	*0*
4	*Join confirmation*	*N*	*0*	*0*	*N*	*N*
5	*LBS relays*	*NT/2*	*0*	*0*	*0*	*0*
6	*Decoy message*	*0*	*0*	*0*	*0*	*NT/2*

Abbreviations: G-2 (GLRP-2), S-R (SMP-R), S-I (SMP-I), O-R (OTFP-R), O-I (OTFP-I)

Table 8.4

Anonymity Metrics and Location Privacy Models

Eq	Metric	G-2	S-R	S-I	O-R	O-I
(1)	$k=\|AS\|$	○	≈	●	●	●
(2)	$H(\|AS\|)$	○	≈	●	●	●
(3)	Pt	≈	≈	≈	≈	≈
(7)	$d=avg(dst)$	○	≈	≈	≈	≈
(12)	$t=\|TS\|$	≈	○	○	●	●

● indicates superior performance compared to other privacy models

○ indicates inferior performance compared to other privacy models

≈ indicates equivalent performance compared to other privacy models

Abbreviations: G-2 (GLRP-2), S-R (SMP-R), S-I (SMP-I), O-R (OTFP-R), O-I (OTFP-I)

malicious collaborators with LBS. OTFP provides the greatest user control as it requires no central authority to coordinate the privacy protocol.

8.5 Performance Evaluation Using Composite Metric, KDT

This study used four methods for evaluating anonymity, anonymity set size (k), distance deviation (d), time of anonymity (t) and number of vehicles anonymized (n). Anonymization points for the various simulation runs were chosen so that n would remain relatively constant. In that fashion we could ascertain the relative differences between the protocols using the other metrics, k, d, and t.

8.5.1 Anonymity Set Size

In the models tested in this study, k is equivalent to anonymity set size, $|AS|$. Both k and $|AS|$ represent the number of vehicles with which a target vehicle may be anonymized. See Figure 8.7.

The models in this study exhibited three noteworthy performance characteristics with respect to anonymity set size. First, as expected, average anonymity set size increased with vehicle density. However, there was a decay effect. In the highest density levels, anonymity did not achieve theoretical performance predictions. Average anonymity set size decreased as large proportions of vehicles exited the region early and became deanonymized. Note: when the very last car in an anonymity set exits the region it has a k value of 1. If large numbers of vehicles exit the region early, the remaining vehicles with lower than predicted average k values deflate average anonymity set sizes. See Figure 8.8.

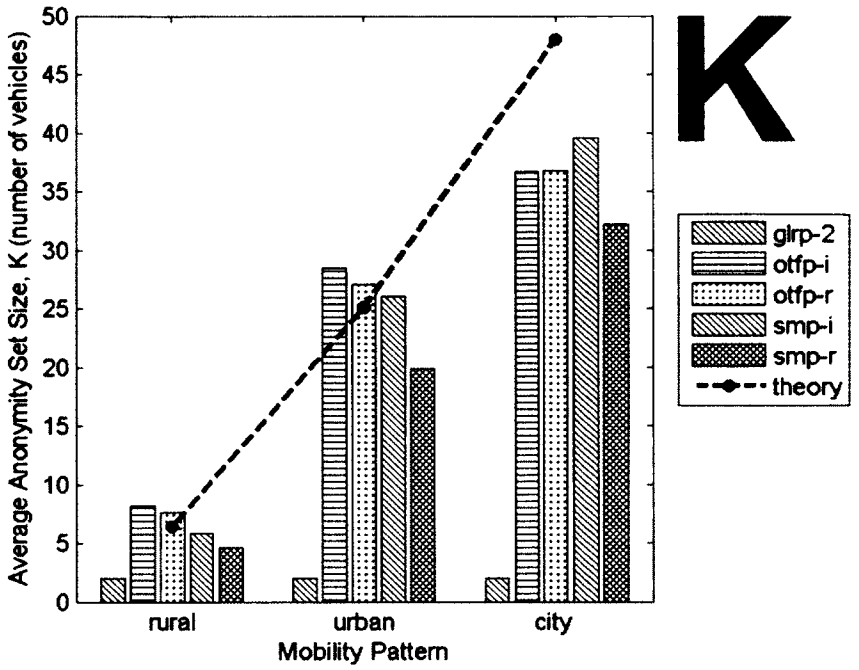


Figure 8.7: Average anonymity set size increased with vehicle density. Theoretical K overestimated simulation at higher vehicle densities. Clumpy privacy protocols (OTFP I, OTFP R and SMP I) provided highest K values.

Second, the "clumpy" models, OTFP-I, OTFP-R and SMP-I, exceeded the performance of the more uniform model, SMP-R. (GLRP-2 exhibited small consistent anonymity set size of 2 because vehicles joined in pairs.) Offsetting the decay effect described in the prior paragraph, which decreases *k* values, clumpiness increases k values. Irregular timing and on-the-fly beacon locations increased clumpiness and improved overall average anonymity. See Figure 8.8.

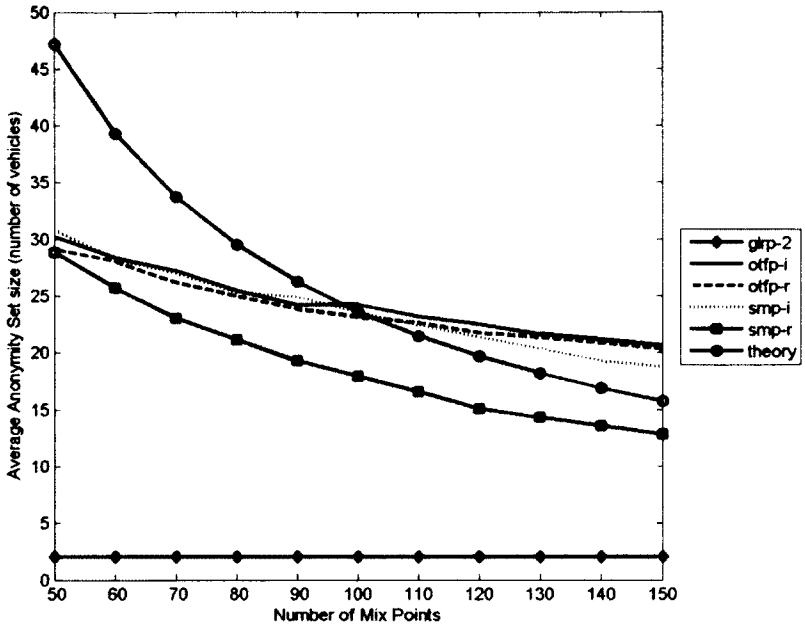


Figure 8.8: Average anonymity set size decreased with number of mix points. SMPrR performed most consistently with theoretical pattern. Clumpy models followed the pattern less closely, retaining higher densities at higher numbers of mix points. Group leader model had fixed average anonymity set size of 2.

Third, the clumpy models were less influenced by the number of mix points. Again, this is because if vehicles are clumped together the number of mix points becomes less relevant. It would be possible for k to equal V, the total number of vehicles, if all vehicles were in the same anonymity set and entered and left the region simultaneously. See Figure 8.9.

Note that the information entropy of the average anonymity set size, $H(|AS|)$, exhibits the same relative characteristics as anonymity set size, $|AS|$. The result differs only in magnitude so is not included here.

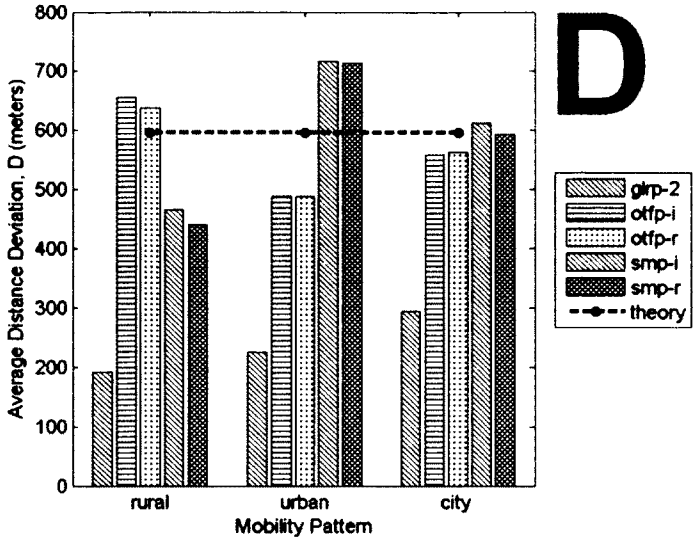


Figure 8.9: Distance deviation fluctuated with vehicle density. The SMP and OTFP models varied in effectiveness at the low, rural, density compared to the medium, urban density. GLRP 2 was lower than all others because vehicles must remain within in communications range (300m) of their group leader at all times.

8.5.2 Distance Deviation

Recall that distance deviation is the average length between a vehicle and all of its peers in the same anonymity set. Roadway layout may affect distance deviation performance. Our tests were confined to the GMSF topologies so this report should be understood in that light.

In low densities the on-the-fly models outperformed other models. In intermediate densities stationary mix points exhibited superior performance. As density increased the models performed equivalently. The most important observation from the analysis of distance deviation is that group leader models perform noticeably worse than other models. This is because vehicles must remain within communications range of the group leader to relay messages. See Figure 8.10.

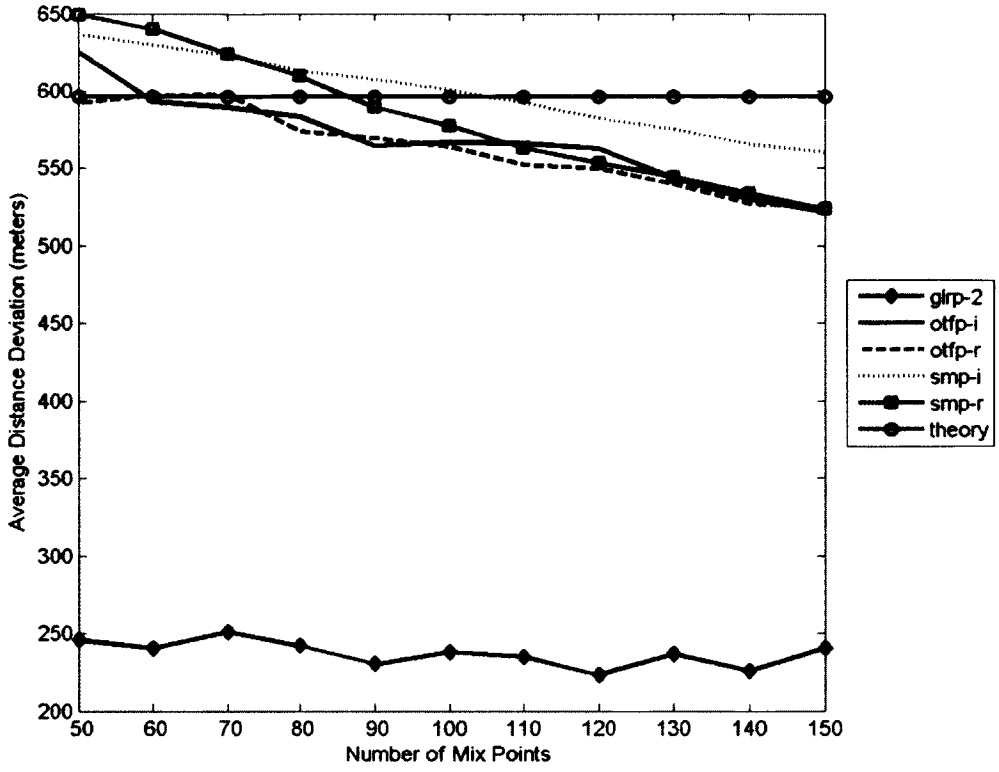


Figure 8.10: Distance deviation decreased with number of mix points. GLRP 2 was lower than all others because vehicles must remain within in communications range (300m) of their group leader at all times.

Theoretical/anticipated average distance deviation closely estimated actual results at high densities. There was an unexpected downward trend when evaluating the average distance deviation prediction given the number of mix points in the simulation. This appears to be due to the initial proximity in the anonymity set. With more anonymity sets (more mix points) each initial group started off closer together which decreased the overall average distance deviation as the number of mix points increased. See Figure 8.11.

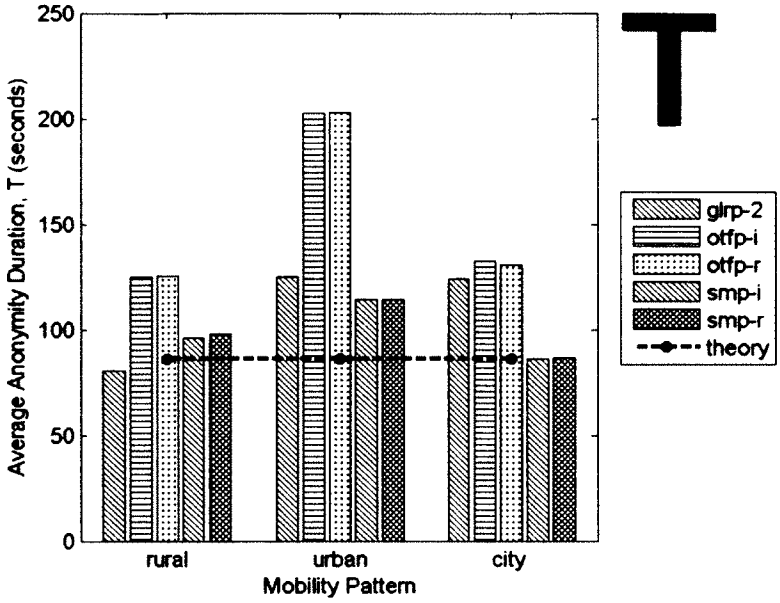


Figure 8.11: Time of anonymity remained consistent with increasing vehicle density. On-the-fly models outperformed other models.

8.5.3 Anonymity Duration

Stationary mix point models performed consistently with predictions regarding the time of anonymity, or anonymity duration of the anonymized vehicles. On-the-fly models exceeded predicted performance at all density levels, and the group model performed better at higher densities. See Figure 8.12.

The reason on-the-fly models performed better is because on average they started closer to the center of the region, the optimal point to maximize time of anonymity. Recall the experiment was set up for vehicles to enter a region, execute a privacy protocol, and then try to maintain anonymity and connectivity before exiting the region, when privacy would no longer be maintained. The group leader model is similar to the on-the-fly model in that it does not use stationary mix points, so GLRP outperformed SMP for the same reason.

In order to maintain constant overall system anonymity, for each density level a stationary mix point was selected based on its ability to anonymize 50% of all vehicles. The region was 3000x3000m so the centroid was (1500,1500). However, for SMP privacy models the rural mobility model set the mix point at (2290,800); the urban mobility model at (1430,2490); and the city mobility model at (390,1710). Stationary mix points being closer to the edge, they were more likely to suffer from vehicles exiting earlier and therefore having shorter times of anonymity, reducing overall time of anonymity.

8.5.4 Number of Vehicles Anonymized

The simulation was set up to ensure relative consistency between models with respect to number of vehicles anonymized, n. Figure 8.12 shows performance results.

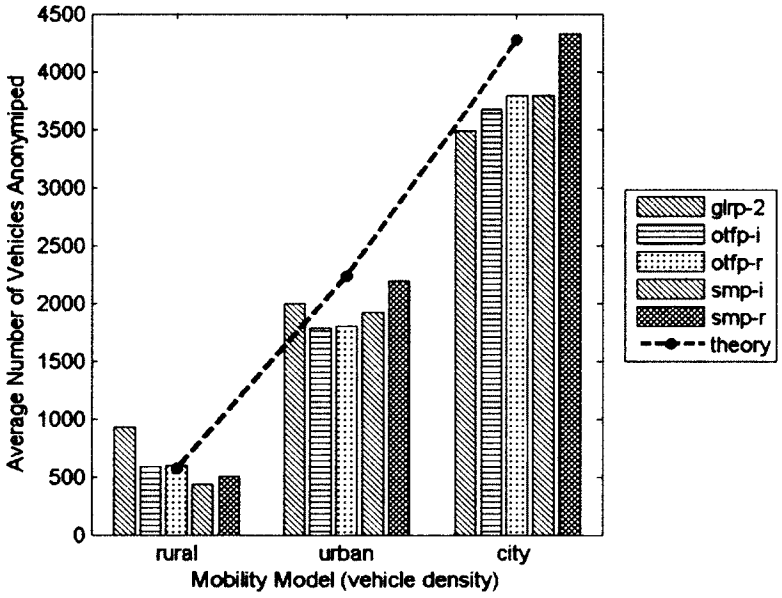


Figure 8.12: Number of vehicles anonymized increased with density. Simulation was set up to ensure relative consistency between models with respect to number of vehicles anonymized.

8.5.5 Summary of Privacy Protocol Performance

Table 8.5 shows privacy protocol performance in terms of KDT anonymity under the urban mobility pattern, specifically, GLRP 2: low low med, OTFP I: high med high, OTFP R: high med high, SMP I: high high low, SMP R: med high low. The new metric identified clear distinctions between the protocols, except for the OTFP protocols. In fact, the new metric helped reveal the effect of clumpiness on privacy, specifically that privacy protocols need only one method of achieving clumpiness. Using two methods, randomness in both timing and positioning, does not improve performance.

8.6 Comparison of KDT with Prior Metrics

This section compares the performance of the KDT metric with prior metrics, trajectory k-anonymity (k), entropy of trajectory k-anonymity (H) and tracking probability (Pt).

8.6.1 K vs. Trajectory k-anonymity

The value of k as defined in (1) is a vehicle's anonymity at the end of its trajectory. So k for the first vehicle to be deanonymized would be k_{max}, the maximum value of the anonymity set size. This assumes the anonymity set size decreases monotonically, as in the simulation. The AS-size of the last vehicle would be 1, assuming vehicles deanonymize one at a time. The average AS-size under these conditions would be $k_{avg} = (1 + 2 + 3 + \ldots + k_{max}) / k_{max}$. The new metric, K, averages AS-size over many time periods, where the old metric considers only the set of trajectories in the anonymity set at the end of the trajectory.

TABLE 8.5

Privacy Protocol Performance (Urban Mobility Pattern)

Protocol	*K*	*D*	*T*	k_{max}	*H*	*Pt*
GLRP-2	2.00	224.40	125.12	2.00	1.00	0.554
OTFP-I	28.42	488.58	202.69	18.81	4.23	0.602
OTFP-R	27.06	487.63	202.93	19.23	4.27	0.597
SMP-I	26.05	717.57	114.08	20.81	4.38	0.571
SMP-R	19.87	714.39	114.16	24.31	4.60	0.509

Why is this important? Because prior to embarking on a trajectory a driver does not know when an attacker will try to deanonymize her. If the attacker investigates historical data and has access to all trajectory information, then the only value of k that matters is the one for the last time interval in the trajectory. However, if a driver wants to choose a privacy protocol that protects her in real time, then it is critical to consider the level of anonymity all along the entire trajectory. This is analogous to the value of car insurance. Looking at the expense historically, if you never had an accident, then car insurance was a waste of money. But looking at the expense as protection against future risks, it may be worthwhile.

Figure 8.13 shows the value of K for each privacy protocol. Figure 8.13 compares the average values of k_{max} (right) with the value of k at the beginning of a trajectory.

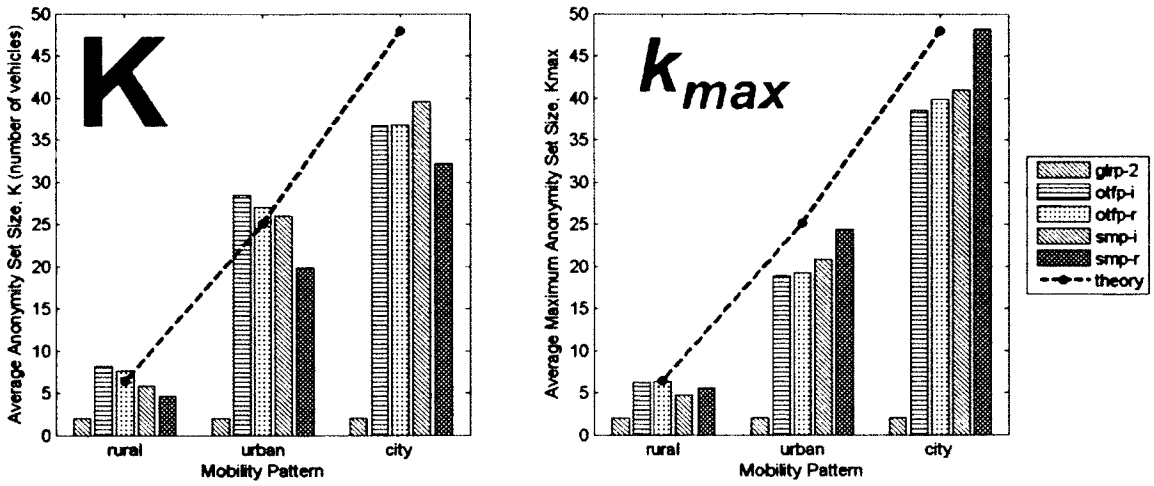


Figure 8.13: K vs. trajectory k-anonymity. The latter does not incorporate the effect of higher anonymity in earlier time intervals and consequently reports lower values of k for clumpy protocols and higher values for uniform protocols. The chart at right shows average maximum anonymity set size, k_{max}.

8.6.2 H[K] vs. Entropy of Trajectory k-anonymity

In evaluating the relative merits of privacy protocols, the same conclusions apply using K vs. k_{max}, as those using $H[K]$ vs. $H[k_{max}]$. See Figure 8.14 and compare with Figure 8.13. Researchers in [25] contend that H is preferable to k to quantify the information content in a trajectory. However, this requires knowing ahead of time how many positions are possible in the mobility pattern. Since we cannot know this for certain we prefer K, rather than $H[K]$.

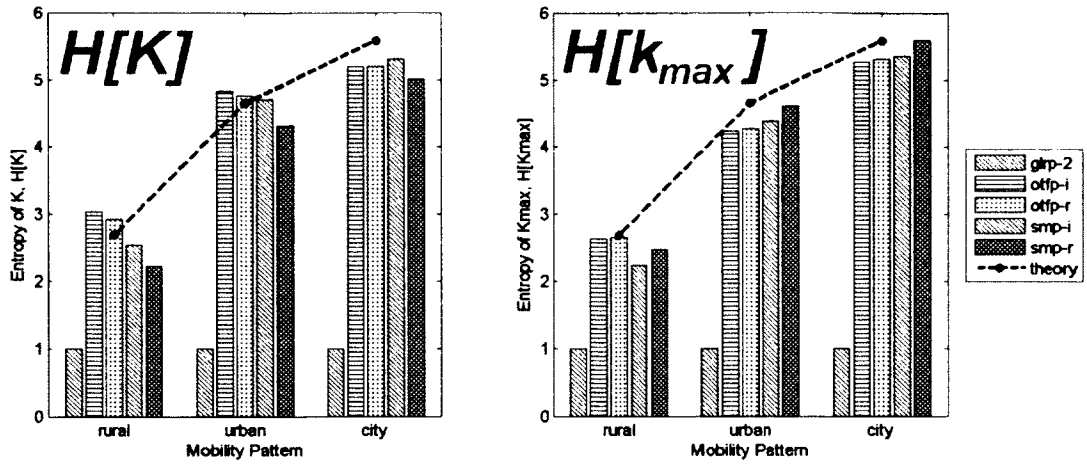


Figure 8.14: Information entropy. The same conclusions regarding privacy protocol performance apply using K vs. k_{max}, as those using $H[K]$ vs. $H[k_{max}]$.

8.6.3 KDT vs. Tracking Probability

Tracking probability is a measure of an entity's chance of having no privacy protection at all. Unlike other metrics, for tracking probability the lower the better. GLRP-2 provided the lowest (best) overall tracking probability, 0.444, a result skewed by performance in rural pattern. All other protocols delivered higher (worse) tracking probabilities than the theoretical prediction, 0.5. See Figure 8.15.

Clumpy protocols performed worse than SMP-R. In general, the higher (better) the average anonymity set size the higher (worse) the tracking probability, but differences were slight and inconsistent. So the decision to use KDT over *Pt* may depend on whether a potential target prefers a high level KDT or a low chance of having no protection at all at rural vehicle densities.

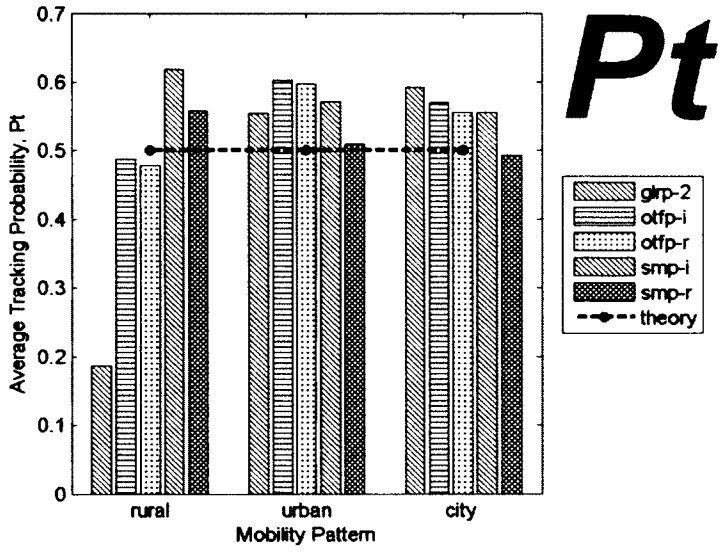


Figure 8.15. Tracking probability. Privacy protocols performed similarly because the simulation was deliberately set up to anonymize roughly half of the vehicles. GLRP-2 provided the lowest (best) overall *Pt* value.

8.7 Summary

To measure continuous network location privacy what is needed is a metric that answers three questions about VANET privacy protocols. (1) How much anonymity does a protocol provide, given trajectories with fluctuating anonymity set sizes? (2) How much dispersion does a protocol provide, given trajectories with fluctuating distance deviations? (3) How long does a protocol anonymize trajectories? This research has proposed KDT anonymity, a composite metric for improving the granularity of the measurement of the property of anonymity over space and time during a trajectory.

Simulation showed how KDT can be used to compare privacy protocols, and how KDT compares with other metrics. The primary advantage of KDT over prior metrics is that it incorporates changes in the magnitude of anonymity over time. While trajectory *k*-anonymity considers AS size at the end of a trajectory, *K* is the average AS size over

the entire duration of the trajectory. Figure 8.15 shows that clumpy privacy protocols have higher K values. This is because more vehicles belong to larger anonymity sets for longer periods of time. This effect is not revealed using trajectory k anonymity.

Another advantage of KDT is it incorporates the magnitude of dispersion, i.e. distance deviation, over time. Under conditions where a potential target values how far she is from her decoys, the D component of KDT becomes a valuable consideration in selecting a privacy protocol.

Finally, if before embarking on a trajectory a potential target values how long she may be anonymized, the T component of KDT becomes a valuable consideration in selecting a privacy protocol.

CHAPTER NINE

CONCLUSION

This dissertation has reviewed dummy event based location privacy protocols, and has offered definitions and metrics which would apply to any location privacy protocols. Specifically, this work has shown the following.

There is a list of twelve properties that help define the scope of desirable attributes of location privacy preserving protocols. This list offers a "big picture" perspective to help categorize and understand past research and perhaps focus current and future research.

A composite metric, KDT, offers advantages over traditional metrics. Vehicles travel fast, often on predictable paths. In order for privacy to be effective, vehicles should be anonymized amongst other vehicles, as in *k*-anonymity, but there should be some significant distance between target vehicles and other members of the anonymity set. Further, there should be significant duration of time during which a vehicle is anonymous. This composite metric would help researchers to compare and contrast privacy protocols over a wider range of important considerations.

Dummy locations which are not realistic in vehicular contexts may be detectable (map deanonymizable) as fakes because location coordinates can be cross-referenced and validated using maps. This problem can be resolved by restricting dummy vehicles to roadways, as in RRVT.

Endpoint protection zones, EPZs, protect vehicle location privacy from deanonymization. If LBS administrators can correlate origin and destination points with

home and work addresses, they can link identity and location of vehicle owners. This is not possible if vehicles remain transmission-silent in their respective EPZs.

EPZs protect vehicle location privacy from collusion between LBS administrators and RSU administrators. Even if LBS administrators can verify with RSUs the locations of transmissions at their points of origination, they cannot be certain which vehicle from the EPZ is making the request unless they have additional information beyond the scope of this study.

The effectiveness of EPZs depends upon density, multiple LBS users originating from each EPZ. In sparsely populated areas, the EPZ model may be ineffective. One workaround might be encouraging local friends and family to use LBS. Another workaround might be to increase EPZ sizes in sparsely populated areas. The practicality of this is a subject of future study. Perhaps the most important finding of this paper is that a small increase in LBS users in a sparsely populated area, which may yield only a proportional effect in anonymity set size, may have a much greater effect on tracking probability.

An interesting property of the EPZ model is that it provides protection even if only one LBS user from an EPZ is active outside that EPZ. The LBS administrator cannot know which LBS user is active. This implies that LBS users especially concerned about privacy could register under multiple false identities, or have friends and family who do not use the LBS register under real identities. In this way a single person could achieve k-anonymity > 1.0.

The EPZ and PARROTS (PBD) protocols address conditions, (1) when the LBS requires continuous precise location data in a vehicular ad hoc network, (2) when the

LBS administrator colludes with administrators of vehicular wireless access points (a.k.a. roadside units, or RSUs), and (3) when precise location data can be deanonymized using map databases linking vehicle positions with vehicle owners' home/work addresses and geographic coordinates. Simulations using realistic vehicle traffic mobility patterns showed PARROTS increased average privacy levels in high vehicle density conditions when deployed in combination with EPZs, and increased them even more effectively in low vehicle density conditions.

An inverse exponential relationship exists between the proportion of vehicles transmitting BSMs and the proportion of potential collisions unprotected by BSMs. If privacy protocols must include a silent period, then perhaps as many vehicles as possible should execute privacy protocols when any vehicles decide to go transmission silent. This is because the marginal loss in collision protection is greatest when one single vehicle decides not to transmit BSMs while all others do. The Safety-Silence Tradeoff Equation (SSTE) can estimate the cost of VANET privacy methods that require silent periods. This can be done by computing the likelihood of crashes between combinations of vehicles at intersections.

Simulation showed how KDT can be used to compare privacy protocols, and how KDT compares with other metrics. The primary advantage of KDT over prior metrics is that it incorporates changes in the magnitude of anonymity over time. While trajectory *k*-anonymity considers AS size at the end of a trajectory, K is the average AS size over the entire duration of the trajectory. Figure 8.15 shows that clumpy privacy protocols have higher K values. This is because more vehicles belong to larger anonymity sets for longer periods of time. This effect is not revealed using trajectory k anonymity.

Another advantage of KDT is it incorporates the magnitude of dispersion, i.e. distance deviation, over time. Under conditions where a potential target values how far she is from her decoys, the D component of KDT becomes a valuable consideration in selecting a privacy protocol.

Finally, if before embarking on a trajectory a potential target values how long she may be anonymized, the T component of KDT becomes a valuable consideration in selecting a privacy protocol.

REFERENCES

[1] Bush, I. (2013, Feb 25). GM, AT&T readying in-vehicle wi-fi. http://philadelphia.cbslocal.com/2013/02/25/gm-att-readying-in-vehicle-wi-fi/

[2] Johnson, L. (2012, Oct 31). Location-based services to bring in $4b revenue in 2012: study. http://www.mobilemarketer.com/cms/news/research/14115.htmlhttp://www.mobilemarketer.com/cms/news/research/14115.html

[3] Koslowski, T. (2012, Jan 3). Your connected vehicle is arriving. http://www.technologyreview.com/news/426523/your-connected-vehicle-is-arriving/

[4] 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) , vol., no., pp.1,289, April 26 2013, doi: 10.1109/IEEESTD.2013.6509896

[5] Kenney, John B. "Dedicated short-range communications (DSRC) standards in the United States." Proceedings of the IEEE 99.7 (2011): 1162-1182.

[6] Corser, G., Fu, H., Shu, T., D'Errico, P., Ma, W. (2013). "Endpoint Protection Zone (EPZ): Protecting LBS User Location Privacy Against Deanonymization and Collusion in Vehicular Networks." The 2nd International Conference on Connected Vehicles & Expo (ICCVE 2013).

[7] Sampigethaya, K., Huang, L., Li, M., Poovendran, R., Matsuura, K., & Sezaki, K. (2005). CARAVAN: Providing location privacy for VANET. Washington Univ Seattle Dept Of Electrical Engineering.

[8] Guo, J., Baugh, J. P., & Wang, S. (2007, May). A group signature based secure and privacy-preserving vehicular communication framework. In 2007 Mobile Networking for Vehicular Environments (pp. 103-108). IEEE.

[9] Sun, Y., Lu, R., Lin, X., Shen, X., & Su, J. (2010). An efficient pseudonymous authentication scheme with strong privacy preservation for vehicular communications. Vehicular Technology, IEEE Transactions on, 59(7), 3589-3603.

[10] Shokri, R., Freudiger, J., & Hubaux, J. P. (2010). A unified framework for location privacy. 3rd Hot Topics in Privacy Enhancing Technologies (HotPETs).

[11] Chow, R., & Golle, P. (2009, November). Faking contextual data for fun, profit, and privacy. In Proceedings of the 8th ACM workshop on Privacy in the electronic society (pp. 105-108). ACM.

[12] Kido, H., Yanagisawa, Y., & Satoh, T. (2005, July). An anonymous communication technique using dummies for location-based services. In Pervasive Services, 2005. ICPS'05. Proceedings. International Conference on (pp. 88-97). IEEE.

[13] Krumm, J. (2009). Realistic driving trips for location privacy. In Pervasive Computing (pp. 25-41). Springer Berlin Heidelberg.

[14] Lu, H., Jensen, C. S., & Yiu, M. L. (2008, June). Pad: Privacy-area aware, dummy-based location privacy in mobile services. In Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access (pp. 16-23). ACM.

[15] You, T. H., Peng, W. C., & Lee, W. C. (2007, May). Protecting moving trajectories with dummies. In Mobile Data Management, 2007 International Conference on (pp. 278-282). IEEE.

[16] Yang, Q., Lim, A., Ruan, X., & Qin, X. (2010, December). Location privacy protection in contention based forwarding for VANETs. In Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE (pp. 1-5). IEEE.

[17] Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), 3.

[18] Li, N., Li, T., & Venkatasubramanian, S. (2007, April). t-closeness: Privacy beyond k-anonymity and l-diversity. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 106-115). IEEE.

[19] Dwork, C. (2006). Differential privacy. In Automata, languages and programming (pp. 1-12). Springer Berlin Heidelberg.

[20] Beresford, A. R., & Stajano, F. (2003). Location privacy in pervasive computing. Pervasive Computing, IEEE, 2(1), 46-55.

[21] Uzcategui, R.; Acosta-Marum, G., "Wave: A tutorial," Communications Magazine, IEEE, vol.47, no.5, pp.126,133, May 2009, doi: 10.1109/MCOM.2009.4939288

[22] Harri, J., Filali, F., & Bonnet, C. (2009). Mobility models for vehicular ad hoc networks: a survey and taxonomy. Communications Surveys & Tutorials, IEEE, 11(4), 19-41.

[23] Baumann, R., Legendre, F., & Sommer, P. (2008, May). Generic mobility simulation framework (GMSF). In Proceedings of the 1st ACM SIGMOBILE workshop on Mobility models (pp. 49-56). ACM.

[24] http://gmsf.sourceforge.net/

[25] http://www.lst.inf.ethz.ch/research/ad-hoc/car-traces

[26] Sweeney, Latanya. "k-anonymity: A model for protecting privacy." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10.05 (2002): 557-570.

[27] Gruteser, M., & Grunwald, D. (2003). Anonymous usage of location-based services through spatial and temporal cloaking. In Proceedings of the 1st international conference on Mobile systems, applications and services (pp. 31-42). ACM.

[28] Corser, G., Fu, H., Shu, T. D'Errico, P., Ma, W. (2013, December). Endpoint Protection Zone (EPZ): Protecting LBS user location privacy against deanonymization and collusion in vehicular networks. In Second International Conference on Connected Vehicles & Expo. IEEE.

[29] Glassbrenner, D. (2003). Estimating the lives saved by safety belts and air bags. US DOT, National Highway Traffic Safety Administration , page, 5, 12.

[30] http://icsw.nhtsa.gov/safercar/ConnectedVehicles/pages/v2v.html.

[31] PBS, Feds to announce decision on new automobile safety technology. Feb 3, 2014. http://www.pbs.org/newshour/rundown/feds-announce-decision-new-automobile-safety-technology/

[32] Dedicated Short Range Communications (DSRC) Message Set Dictionary, SAE Std J2735, SAE International,DSRC committee, 2009. SAE.

[33] Raya, M., & Hubaux, J. P. (2005, September). The security of VANETs. In Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks (pp. 93-94). ACM.

[34] TS3290/00A. On-Board Unit. Kapsch. Accessed 2014-04-23. https://www.kapsch.net/ktc/downloads/datasheets/in-vehicle/5-8/Kapsch-KTC-DS-OBU-TS3290_00A?lang=en-US

[35] Lee, E. K., Yang, S., Oh, S. Y., & Gerla, M. (2009, October). RF-GPS: RFID assisted localization in VANETs. In Mobile Adhoc and Sensor Systems, 2009. MASS'09. IEEE 6th International Conference on (pp. 621-626). IEEE.

[36] Bohlooli, A., & Jamshidi, K. (2012). A GPS-free method for vehicle future movement directions prediction using SOM for VANET. Applied Intelligence, 36(3), 685-697.

[37] http://www-fars.nhtsa.dot.gov/Main/index.aspx

[38] http://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm

[39] Haas, J. J., Hu, Y. C., & Laberteaux, K. P. (2011). Efficient certificate revocation list organization and distribution. Selected Areas in Communications, IEEE Journal on, 29(3), 595-604.

[40] http://www.merriam-webster.com/dictionary/anonymous

[41] http://www.merriam-webster.com/dictionary/pseudonymous

[42] Diaz, C., Claessens, J., Seys, S., & Preneel, B. (2002, May). Information theory and anonymity. In Proceedings of the 23rd Symposium on Information Theory in the Benelux (pp. 179-186).

[43] Shokri, R., Troncoso, C., Diaz, C., Freudiger, J., & Hubaux, J. P. (2010, October). Unraveling an old cloak: k-anonymity for location privacy. In Proceedings of the 9th annual ACM workshop on Privacy in the electronic society (pp. 115-118). ACM.

[44] Coull, S. E., Wright, C. V., Keromytis, A. D., Monrose, F., & Reiter, M. K. (2008). Taming the devil: Techniques for evaluating anonymized network data. In Network and Distributed System Security Symposium 2008: February 10-13, 2008, San Diego, California: Proceedings (pp. 125-135). Internet Society.

[45] Xiao, X., & Tao, Y. (2007, June). M-invariance: towards privacy preserving re-publication of dynamic datasets. In Proceedings of the 2007 ACM SIGMOD international conference on Management of data (pp. 689-700). ACM.

[46] Dewri, R., Ray, I., & Whitley, D. (2010, May). Query m-invariance: Preventing query disclosures in continuous location-based services. In Mobile Data Management (MDM), 2010 Eleventh International Conference on (pp. 95-104). IEEE.

[47] Kelly, D. J., Raines, R. A., Grimaila, M. R., Baldwin, R. O., & Mullins, B. E. (2008, October). A survey of state-of-the-art in anonymity metrics. In Proceedings of the 1st ACM workshop on Network data anonymization (pp. 31-40). ACM.

[48] Park, H., & Kent, S. (2009). Traceable anonymous certificate.

[49] Chaum, D. (1983, January). Blind signatures for untraceable payments. In Advances in cryptology (pp. 199-203). Springer US.

[50] Corser, G., Fu, H., Shu, T. D'Errico, P., Ma, W., Leng, S., Zhu, Y. (2014, June). Privacy-by-Decoy: Protecting Location Privacy Against Collusion and Deanonymization in Vehicular Location Based Services. In 2014 IEEE Intelligent Vehicles Symposium. IEEE. Dearborn, MI.

[51] Alnahash, N., Corser, G., Fu, H. (2014, April). Protecting Vehicle Privacy using Dummy Events. In 2014 American Society For Engineering Education North Central Section Conference (ASEE NCS 2014).

[52] M. Raya and J.-P. Hubaux, "Security Aspects of Inter-Vehicle Communications," *in Proc. of Swiss Transport Research Conference*, March 2005.

[53] M. Raya and J.-P. Hubaux, "The Security of Vehicular Ad Hoc Neworks," *in Proc. of the ACM Workshop on Security of Ad hoc and Sensor Networks (SASN)*, pp. 11– 21, November 2005, pp. 11–21.

[54] F. Dotzer, "Privacy Issues in Vehicular Ad hoc Networks," *in Proc. of the Workshop on Privacy Enhancing Technologies (PET)*, June 2005, pp. 197–209.

[55] E. Schoch, F. Kargl, T. Leinmuller, S. Schlott, and P. Papadimitratos, "Impact of Pseudonym Changes on Geographic Routing in VANETs," *in Proc. of the European Workshop on Security and Privacy in Ad hoc and Sensor Networks (ESAS)*, October 2006, pp. 43–57.

[56] L. Buttyan, T. Holczer, and I. Vajda, "On The Effectiveness Of Changing Pseudonyms to Provide Location Privacy in Vanets," *InProc of European workshop on security and privacy in ad hoc and sensor networks (ESAS)*, Cambridge, 2007.

[57] J. J. Haas, Y.-C. Hu, and K. P. Laberteaux, "The Impact of Key Assignment on VANET Privacy," *Security and Communication Networks 2009*, John Wiley & Sons, Ltd., September, 2009.

[58] Y. Sun, R. Lu, X. Lin, X. Shen, and J. Su, "NEHCM: A Novel and Efficient Hash-chain based Certificate Management Scheme for Vehicular Communications," *Proc. Chinacom'10*, Beijing, China, August 25-27, 2010.

[59] L. Zhang, Q. Wu, A. Solanas, and J. Domingo-Ferrer, "A Scalable Robust Authentication Protocol for Secure Vehicular Communications," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1606-1617, May 2010.

[60] Y. Sun, B. Zhao, and J. Su, "ECHO: Efficient Certificate Updating Scheme by Vehicle-to-Vehicle Communications," *The 4th International Conference on Frontier of Computer Science and Technology (FCST 2009)*, Shanghai, China, December 17-19, 2009.

[61] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos, "Privacy In Inter-Vehicular Networks: Why Simple Pseudonym Change Is Not Enough," *in Proc. 7th International Conference on wireless On-demand Network Systems and Services (WONS'10)*, pp. 176-183, 2010.

[62] Y. Sun, R. Lu, X. Lin, X. Shen, and J. Su, "An Efficient Pseudonymous Authentication Scheme with Strong Privacy Preservation for Vehicular Communications," *IEEE Trans. on Vehicular Technology*, to appear.

[63] J. J. Haas, Y.-C. Hu, and K. P. Laberteaux, "Design and Analysis of a Lightweight Certificate Revocation Mechanism for VANET," *in Proceedings of the Sixth ACM International Workshop on Vehicular Ad Hoc Networks (VANET 2009)*. ACM, Beijing, China, September 2009.

[64] M. Gruteser and D. Grunwald, "Anonymous Usage Of Location-Based Services Through Spatial And Temporal Cloaking," *in Proc. of the ACM International Conference on Mobile Systems MobiSys*, May 2003, pp. 31–42.

[65] S. Kato, S. Tsugawa, K. Tokuda, T. Matsui, and H. Fujii, "Vehicle Control Algorithms For Cooperative Driving With Automated Vehicles And Intervehicle Communications," *IEEE Trans. on Intelligent Transportation Systems*, vol. 3, no. 3, pp. 155–161, September 2002.

[66] R. Hochnadel and M. Gaeta, "A Look Ahead Network (LANET) Model for Vehicle-to-vehicle Communications Using DSRC," *in Proc. of World Congress on Intelligent Transportation Systems*, November 2003.

[67] M. Raya, A. Aziz, and J.-P. Hubaux, "Efficient Secure Aggregation in VANETs," *in Proc. of the 3rd international workshop on Vehicular Ad hoc Networks (VANET)*, pp. 67–75, 2006.

[68] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, "Enhancing Wireless Location Privacy Using Silent Period," *in Proc. of the IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1187–1192, March 2005.

[69] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, "Towards Modeling Wireless Location Privacy," *in Proc. of the Workshop on Privacy Enhancing Technologies (PET)*, pp. 59–77, June 2005.

[70] D. Chaum, "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability," *Journal of Cryptology*, vol. 1, pp. 65–75, 1988.

[71] Serjantov and G. Danezis, "Towards an Information Theoretic Metric for Anonymity," *in Proc. of the Workshop on Privacy Enhancing Technologies (PET)*, pp. 41–53, April 2002.

[72] F. L. Mannering, W. P. Kilareski,and S. S. Washburn, "Principles of Highway Engineering and Traffic Analysis," 3rd Ed, ISBN: 978-0-471-47256-8, Wiley Publishers, 2004.

[73] D. Boneh, X. Boyen, and H. Shacham, "Short Group Signature," *in Proc. Advances in Cryptography – Crypto'04, ser. LNCS, vol. 3152, Springer-Verlag*, pp. 41-55, 2004.

[74] J. H. Song, V. W. S. Wong, and V. C. M, Leung, "Wireless Location Privacy Protection in Vehicular Ad hoc Networks," *Springer-Mobile Networks and Applications*, Vol. 15, No. 1. , pp. 160-171, Feb. 2010.

[75] G. Yvonne, W. Bernhard, and G. Hans Peter, "Medium Access Concept for VANETs Based on Clustering," *in the 66th IEEE VTC*, pp. 2189-2193, Sep. 2007.

[76] P. Fan, "Improving Broadcasting Performance by Clustering with Stability for Inter-vehicle Communication," *Proceedings of the 65th IEEE VTC*, Dublin, Ireland, April 2007.

[77] W. Zhiagang, L. Lichuan, Z. MengChu, and A. Nirwan, "A Position-Based Clustering Technique for Ad Hoc Intervehicle Communication," *IEEE transactions on Man and Cybernetics*, vol.38, No.2, March 2008.

[78] Z. Y. Rawashdeh and S. M. Mahmud, "Toward Strongly Connected Clustering Structure in Vehicular Ad Hoc Networks," *Proceedings of the 70th IEEE VTC*, Alaska, USA, Sep. 2009.

[79] L. Buttyán, T. Holczer, A. Weimerskirch, and W. Whyte, "SLOW: A Practical Pseudonym Changing Scheme for Location Privacy in VANETs," *IEEE Vehicular Networking Conference (VNC)*, Tokyo, Japan, October 2009.

[80] IEEE P802.11p TM/D3.0 Draft Standard for Information Technology – Telecommunications and Information Exchange between Systems - Local and metropolitan area networks - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.

[81] Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos, "Privacy in Inter-Vehicular Networks: Why Simple Pseudonym Change is Not Enough," *in Proc. 7th International Conference on wireless On-demand Network Systems and Services (WONS'10)*, pp. 176-183, 2010.

[82] Gedik and L. Lu, "Location Privacy in Mobile Systems: A Personalized Anonymization Model," *In Proceedings of the 25th IEEE ICDCS 2005*, pp 620-629, Washington, DC, USA, 2005.

[83] B. Hoh and M. Gruteser, "Protecting Location Privacy Through Path Confusion," *In Proceedings of IEEE Create-Net Secure Comm*, Athens, Greece, Sep. 2005.

[84] L. Zhou and Z. J. Haas, "Securing Ad Hoc Networks," *IEEE Net.*, vol. 13, no. 6, Nov./Dec. 1999.

[85] J. Lundberg, "Routing Security in Ad Hoc Networks," Helsinki University of Technology, http://citeseer.nj.nec.com/400961.html

[86] M. Royer and C. K. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," *IEEE Personal Communications*, pp. 46-55, April 1999.

[87] B. Johnson and D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," *Mobile Computing*, edited by Tomasz Imielinski and Hank Korth, Chapter 5, pages 153-181, Kluwer Academic Publishers, 1996.

[88] C. E. Perkins and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers," *Computer Communications Review*, pp. 234-244, October 1994.

[89] H. Deng, W. Li, and D. P. Agrawal, "Routing Security in Wireless Ad Hoc Network," *IEEE Communications Magzine*, vol. 40, no. 10, October 2002.

[90] Sanjay Ramaswamy, Huirong Fu, Manohar Sreekantaradhya, John Dixon, and Kendall Nygard, "Prevention of Cooperative Black Hole Attack in Wireless Ad Hoc Networks," *International Conference on Wireless Networks (ICWN'03)*, Las Vegas, Nevada, USA, June 2003.

[91] Jian Yin and Sanjay Madria, "A Hierarchical Secure Routing Protocol against Black Hole," *IEEE SUTC 2006* Taiwan, 5-7 June 2006.

[92] Xiaoyan Hong, Kaixin Xu, and Mario Gerla, "Scalable Routing Protocols for Mobile Ad hoc Networks," *IEEE Network* Vol.16(4), pp11-21, July/August 2002.

[93] Elizabeth M. Royer and Chai-Keong Toh, "A Review of Current routing Protocols for Ad Hoc Mobile Wireless Networks," *IEEE Personal Communications*, pp 46-55, April 1999.

[94] Sanjay Ramaswamy, Huirong Fu, and Kendall E. Nygard, "Simulation Study of Multiple Black Holes Attack on Mobile Ad Hoc Networks," *International Conference on Wireless Networks (ICWN' 05)*, Las Vegas, Nevada, Jun. 2005.

[95] Hesiri Weerasinghe and Huirong Fu, "ESAP: Efficient and Scalable Authentication Protocol for Vehicular Ad hoc Networks," *IEEE Globecom 2010 Workshop on Ubiquitous Computing and Networks (UbiCoNet 2010)*, December, 2010.

[96] Hesiri Weerasinghe and Huirong Fu, "Enhancing Location Privacy for Vehicular Ad hoc Networks," under Preparation

[97] Hesiri Weerasinghe, Huirong Fu, and Supeng Leng, "Anonymous Service Access Protocol for Vehicular Ad hoc Networks," *6th International Conference on Information Assurance and Security (IEEE IAS 2010)*, Atlanta, USA, August, 2010.

[98] Hesiri Weerasinghe, Huirong Fu, and Raymond Tackett, "Verifying Position and Velocity for Vehicular Ad-Hoc Networks," *Wiley Journal of Security and Communication Networks*, 15 July 2010.

[99] Huirong Fu, Hesiri Weerasinghe, and Ye Zhu, "Improving Hierarchical Mobile Certification Authority for Wireless Ad Hoc Networks," Submitted to Elsevier Journal of Computer Communication

[100] Hesiri Weerasinghe and Huirong Fu, "Preventing Cooperative Black Hole Attacks in Mobile Ad Hoc Networks: Simulation Implementation and Evaluation," *IJSEIA – International Journal of Software Engineering and Its Applications*, vol. 2, No. 3, July 2008.

[101] Hesiri Weerasinghe and Huirong Fu, "Preventing Cooperative Black Hole Attacks in Mobile Ad Hoc Networks: Simulation Implementation and Evaluation," *FGCN WAMSNet-07*, South Korea, Dec. 6-8 2007.

[102] Hesiri Weerasinghe, Huirong Fu, and Supeng Leng, "Providing Location Privacy for Online-services in Vehicular Ad hoc Networks", accepted for *Journal of Information Assurance and Security.*