MODELING ADVERSARIAL INSIDER VEHICLES IN MIX ZONES

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Nicholas Plewtong

March 2018

COMMITTEE MEMBERSHIP

TITLE:	Modeling Adversarial Insider Vehicles in Mix Zones
AUTHOR:	Nicholas Plewtong
DATE SUBMITTED:	March 2018
COMMITTEE CHAIR:	Bruce DeBruhl, Ph.D. Assistant Professor of Computer Science
COMMITTEE MEMBER:	Theresa Migler, Ph.D. Lecturer of Computer Science
COMMITTEE MEMBER:	Anthony Mendes, Ph.D. Professor of Mathematics

ABSTRACT

Modeling Adversarial Insider Vehicles in Mix Zones

Nicholas Plewtong

Security is a necessity when dealing with new forms of technology that may not have been analyzed from a security perspective. One of the latest growing technological advances are Vehicular Ad-Hoc Networks (VANETs). VANETs allow vehicles to communicate information to each other wirelessly which allows for an increase in safety and efficiency for vehicles. However, with this new type of computerized system comes the need to maintain security on top of it.

In order to try to protect location privacy of the vehicles in the system, vehicles change pseudonyms or identifiers at areas known as mix zones. This thesis implements a model that characterizes the attack surface of an adversarial insider vehicle inside of a VANET. This adversarial vehicle model describes the interactions and effects that an attacker vehicle can have on mix zones in order to lower the overall location privacy of the system and remain undetected to defenders in the network. In order to reach the final simulation of the model, several underlying models had to be developed around the interactions of defender and attacker vehicles.

The evaluation of this model shows that there are significant impacts that internal attacker vehicles can have on location privacy within mix zones. From the created simulations, the results show that having one to five optimal attackers shows a decrease of 0.6%-2.6% on the location privacy of the network and a 12% decrease in potential location privacy in a mix zone where an attacker defects in a 50-node network. The industry needs to consider implementing defenses based on this particular attack surface discussed.

ACKNOWLEDGMENTS

Thanks to:

- Dr. DeBruhl for being a great thesis advisor and for all the guidance he has given me.
- Dr. Migler and Dr. Mendes for serving on my committee.
- My Mom for being my number one supporter through everything I do.
- My Dad for pushing me towards computer science.
- My girlfriend, Tiffany, for always being there for me.
- Leanne Fiorentino for being a huge help to me for anything regarding the Computer Science department
- The entire Cal Poly Computer Science Department
- Andrew Guenther, for uploading this template

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Over the last few years, there has been an increasing interest in developing more technological advances in intelligent transportation systems. In particular, much research has gone into Vehicular Ad-Hoc Networks (VANETs) to improve efficiency and safety of transportation by enabling vehicles to communicate between each other [6][8][18]. The applications for VANETs range from safety systems, like collision detection systems, to those that improve efficiency, such as platooning. The use of VANETs will become in industry as the automotive industry develops fully autonomous vehicles.

However, as we seek to create innovative applications for VANETs, the concern of the implementation of security within these applications is worth heeding [15]. One of the main security problems that exists in VANETs is maintaining a vehicle's location privacy [2][3][11][16]. Vehicle nodes differ from mobile and wifi nodes in the way that they not only communicate with other vehicle nodes but also obtain sensor data about their environment, obtain GPS data, and communicate these to both vehicle nodes and roadside nodes [16]. These additional operations and the necessity to work in a high speed environment produce complications in securing the network. An example of this problem is that because the way vehicles broadcast their communications to other vehicles and units, an adversary can eavesdrop on the messages sent from a vehicle. This effectively allows for an adversary to track the location of a vehicle overtime leading to the loss of privacy of that node.

One proposed approach is to combat an adversaries trying to track the location of vehicles by pseudonym switching [12]. Instead of a vehicle using a single identifier for the entire time, the vehicle will periodically change its pseudonym with other

vehicles in order to mitigate an adversary's tracking capabilities. These areas where vehicles switch pseudonyms are known as mix zones [3]. When all vehicles cooperate in a mix zone, then the location privacy of the system increases. However, in terms of self-maintained location privacy, vehicle nodes may decide to not cooperate in self interest since pseudonym changes are costly in terms of resources. If the costs of cooperating in the mix zone outweigh the benefits of location privacy gained for the node, the selfish node would choose to not cooperate. The high density of vehicles as opposed to mobile nodes in a network make their interactions in mix zones significant in maintaining location privacy.

Although previous models have developed on a system of selfish vehicle nodes that act in the interest of maximizing their own location privacy, from security perspective, there can be other type of vehicle nodes with different goals. In this work, I develop a different type of vehicle node acting as a malicious adversary. The goal of the malicious adversary is to minimize the location privacy of the entire system of nodes rather than try to maximize its own location privacy. A core aspect of examining this system of player in the vehicle system of nodes is to demonstrate the effect of a malicious adversary often found in many different types of computer-controlled environments.

1.1 Problems Within Current Models

While current models do examine the important aspect that involves vehicle nodes acting in a selfish manner, it is important that the possibility of nodes with other overall goals is taken into account [13]. If models are created without taking all possibilities of node behaviors into account, then the models may not mimic what behaviors could be found in real VANETs. When talking about a common vehicle node's goal, this corresponds to maintaining of an individual node's security, which, in

this case, is the node's location privacy. In computer security environments, there are usually adversaries that are a counterpart to the goals of the common user, achieving malicious activity such as forcing denial of the services of a system or obtaining secretive data from the system or users. VANETs are a type of computerized network that requires security in order to protect vehicles in the network. Considering that security is required in VANETs, it is possible to consider the goal of an adversary in this type of system. One possible goal would be to lessen the overall security of the network as a whole. This type of adversary has not to the best of the author's knowledge been considered.

1.2 Thesis Overview

In this thesis, I develop an in-depth model and evaluation of how an internal attacker can affect location privacy of a system containing greedy nodes. This is achieved by evaluating the interaction between defender nodes wanting to maximize their own individual location privacy and attackers wanting to minimize the system's overall location privacy through cooperating or defecting in a mix zone.

The contribution of this thesis is to develop and evaluate a model for an internal unknown attacker in a system of greedy vehicle nodes that has not been previously developed. It expands on previous game theoretic mix zone papers to introduce a new type of player as an attacker in the system. The final model shows results of the attackers who not only want to accomplish lowering the location privacy of the system but also wish to remain undetected.

The main contributions from the thesis are:

- Creation of a formal game model for the attack and defense scenario.
- Creation of a model for suspicion levels for defenders to identify potential at-

tackers.

- In a particular 50-node network simulation, development of the optimal strategy for an adversarial insider vehicle as a 37% defect rate.
- In a particular 50-node network simulation, calculation of a 0.6%-2.6% decrease of location privacy by attackers overall on a network.
- In a particular 50-node network simulation, calculation of a 12% decrease of potential location privacy in a mix zone where an attacker defects.

These results help show the effectiveness of this type of attacker in VANETs and thus implies the needs for future research for defenses based on this type of attacker to be modeled for VANETs.

1.3 Outline of Thesis

In chapter 2, the background of security, VANETs, pseudonym switching and game theory are explained. In chapter 3, related works to the content of the thesis are discussed. In chapter 4, the design of the attack and defend model are evaluated. In chapter 5, the implementation of the model and results of the models and simulations are described in detail. In chapter 6, significant findings from our results are discussed. In chapter 7, future work that could be expand from the results of the thesis further discussed. In chapter 8, a summary of the contributions of the thesis is presented.

Chapter 2

BACKGROUND

In order to accurately start creating an internal adversarial model for pseudonym switching in VANETs, this background chapter describes the technical terms related to security, VANETs, pseudonym switching, and game theory. This background provides basic concepts at a general level to better understand subsequent application and contribution of the model.

2.1 Security

Security is the main focus this thesis contributes to. When technological advances develop in computer systems, security is often the last aspect to be considered in comparison to functionality of the system. Yet, if security of a system is not properly maintained, the system's functionality as a whole is threatened to collapse due to possible damage and attacks. As a result, it is important to maintain proper security as well as functionality for a service to work properly.

Security is a mechanism that protects computer systems from attacked while maintaining confidentiality, integrity, and availability of the system. When any of these aspects is compromised and fails to be properly maintained, the security of the system as a whole is jeopardized. It requires all three aspects to be upheld at all times.

2.1.1 Confidentiality and Privacy

Confidentiality is defined as making sure that information and data does not end up in the hands of unauthorized individuals. This means that to protect confidential-

ity, there must be defenses and countermeasures to ensure that data and information is only seen by those authorized to access. This is the aspect that will be most focused on when talking about how to secure VANETs.

One of the main problems in VANETs that threatens the security of the system is the possibility of tracking a location of a vehicle over time. When speaking in terms of confidentiality and location tracking, a vehicle's location over time is confidential information that should only be available to the corresponding vehicle itself. If an unauthorized entity were to be able track another vehicle's location, this breaks the confidentiality and privacy of the vehicle and thus threatens the security of the VANET [9]. Security researchers want to make sure that attack vectors on confidentiality are properly modeled in order to adequately model the defenses necessary to develop to prevent fruitful attacks.

2.1.2 Vulnerabilities

Another concept that pertains to the security of VANETs is the concept of vulnerabilities. A vulnerability is where there exists a weaknesses in the system that enables an attacker to exploit them. In terms of VANETs, attackers seek vulnerabilities in the system in order to disrupt the confidentiality, integrity, and availability of the network.

Since VANETs rely heavily on communications to exchange information between nodes in the network, attackers specifically look for vulnerabilities within the process of the vehicles communicating. This can vary from trying to access information from the communication, modify information being transferred between nodes or even understanding the flaws in the defenses implemented in other attacks. Overall, to understand how to protect the system, research needs to be conducted on what weaknesses currently exist on the network so that appropriate defenses can be implemented.

2.2 Vehicular Ad-Hoc Networks

This section talks about concept of Vehicular Ad-Hoc Networks (VANETs) in order to understand the system that needs to be secured. Vehicular networks are an expansion of a mobile ad hoc network which is a type of network in which the nodes move around and change its locations [7]. Along with this, each node in the network is able to communicate wirelessly with other nodes in the network to transmit information. For VANETs in particular, each node specifically represents a vehicle in the system, which can communicate wireless with other vehicles that exist in the network.

One of the emerging research areas in which VANETs research has been focused on is the concept of autonomous vehicles. With a shift in the capabilities of technology and computer systems growing over the last decade, automotive companies are moving away from human driven vehicles where the human driver is in complete control of all actions of the vehicle to computerized vehicles. With the push towards fully autonomous vehicles, VANETs research has pushed towards other autonomous applications within vehicles being developed.

2.2.1 Communication Model

For vehicle nodes in a VANET, vehicles are able to communicate to other vehicles and roadside infrastructure through different types of computerized systems on board the vehicle. The primary communication system used by the vehicles is dedicated short-range communication (DSRC). This type of communication is used to transmit data quickly between vehicles in order to facilitate the process of the data for applications [8]. DSRC has been developed for securing message authentication and privacy while also making sure that the moving vehicles in the network are able to

identify each other and transmit messages quickly and accurately. It also attempts to mitigate the effect of extra noise throughout the network caused by outside broadcasts and weather conditions.

One the main security concerns involving DSRC stems from the broadcasting nature of the communication. Because vehicles broadcast their information and data for other vehicles in the network to use, attackers can attempt to obtain this data by eavesdropping on the network through using listening nodes [11]. These eavesdropping nodes can be placed all along the route of the network in order to obtain information about the vehicles at different points of time. Because of this, a vehicle's privacy is threatened if an attacker is able to obtain identifying information by tracking a vehicle along the route it is taking.

2.2.2 Applications inside VANETs

The innovation and research that has gone into VANETs has led to vast improvements in the capabilities for safety and efficiency for vehicles. Since vehicles can use DSRC and other sensor equipment to obtain data on the current environment and transfer that information around the network, vehicles can process that data to use for cooperation processes. Platooning is a driver assisting technology that allows vehicles in a system to speed up and slow down in synchronization with other vehicles while maintaining safe spacing between the vehicles [8]. Other benefits to the safety and efficiency of vehicles via having computerized systems from VANETs are collision avoidance, lane keep assistance and traffic optimization [17].

All of these improvements are a result of the innovation in the field of VANETs. Yet, there has not been adequate research in the security involved in securing these processes. Security is usually the last thing in mind in these researches when it comes to developing the processes. That's why security research needs to be improved while

these processes are being developed because of the critical nature of VANETs. It is necessary to secure attacks against the system that could ultimately affect the lives of drivers and passengers. If an attack affects a vehicle's broadcast to other vehicles about speed or location, that could lead to misleading decisions by the vehicle's computer system.

2.3 Pseudonym Switching

The protection of the location privacy, the ability to prevent others from learning one's current or past location [4]. of users is important in VANETs. When vehicles share information via communication channels, they have a unique identifier that allows other vehicles in the system to identify which vehicle the data is coming from. However, if this identifier were to remain constant, other outside users may be able to de-anonymize users based on broadcasted messages corresponding to the constant identifier. In terms of location privacy, if an adversary were to detect that the same identifier was being broadcasted along a route, the adversary would essentially be able to track the vehicle's location and path over a period of time. This loss of privacy is something that needs to be prevented in the way VANETs communicate.

2.3.1 Mix Zones

Because of the troubles brought by a constant identifier in broadcasts, researchers have created a new solution where users mix or switch their identifiers in a temporary zone known as a mix zone. The result afterwards is that vehicles are now communicating with a different identifier than their previous one [2]. This changing identifier is known as a pseudonym. When a vehicle is manufactured, the vehicle contains a set of pseudonyms from which it can switch from.

The general model of a mix zone is as follows [3]:

1. There is a trusted middleware system for the users.

2. Users enter regions known as the mix zone where location can not be tracked.

3. When users decide to cooperate in a mix zone to switch identifiers, the users' identifiers are switched along with all other cooperating users in the zone.

4. When users exit the mix zone region, they can resume communication with their new pseudonym.

The goal of the mix zone is to prevent long term tracking users' location but still allowing an individual short term location application to work. The internal trusted system is able to correlate the pseudonyms to the correct user identity while non-trusted users are able to see the changing pseudonyms that are broadcasted in messages.

2.3.2 Effectiveness of Pseudonym Changing Schemes

One of the key areas of research when it comes to pseudonyms is the effectiveness of the changing scheme. In many cases, this is evaluated by how likely is an adversary able to connect changes in pseudonyms to a corresponding user. If an attacker is able to identify the connection, then the pseudonym changing scheme is not secure. While the specifics on the type of pseudonym-changing scheme is not the focus of this thesis, many past papers have evaluated different types of schemes and their effectiveness. One of the important concepts that will be considered in this thesis is the concept of level of privacy gained. This considers how much location privacy an individual user gains from entering and exiting a mix zone. It is important to have a secure pseudonym changing-scheme in order to further evaluate the severity of other attack threat models that could effect the location privacy of the system.

2.4 Game Theory

Mix zones provide a huge benefit in protecting the location privacy of a system of vehicle nodes when all the nodes cooperate in switching pseudonyms when they enter the mix zones. The flaw in this is that the system is often analyzed only under the assumption that nodes always cooperate. There is, however, rational reasoning for a vehicle to not cooperate and instead defect from switching pseudonyms. This conflicting idea of cooperation versus non-cooperation is the reason why researchers have started to model the interactions in mix zones through game theory.

Game theory is a study of mathematics that deals with modeling interactions of players who make decisions that affect other players in the system. Each player has to consider other players' strategies in order to analyze what their own strategy should be. Game theory contributes a different perspective on how to model the concept of mix zones. Each vehicle in the system now has to analyze whether or not other vehicles will cooperate in switching pseudonyms in order to determine whether or not it should decide to switch. Although this thesis does not go into specific mathematical proofs of game theory, it does apply key aspects of game theory into the consideration of modeling how an attacker would interact with a group of selfish vehicle nodes and show the effect of the attacker's interactions. The following terms are key concepts from game theory that are considered throughout this thesis.

2.4.1 Players

Players of a game are essentially a decision maker in a game. Players of the same type will be modeled in the same way in terms of possible strategies and payoffs gained. The number of players in a game is not bound and provide the basis of what strategies and payoffs need to be defined in a game.

In terms of mix zones, players are individual vehicles that exist in the system. These vehicles will have a defined strategy model and payoffs that they can gain in the game which is the interaction in the mix zone. Different types of vehicles as players can be defined and discussed later in this thesis.

2.4.2 Strategy

Strategy is one of the basic concepts that exist in game theory when describing a game. It is the one of the options from which a player is able to choose. This strategy not only affects the result for that player, but also affects the results for all the other players in the game. Another property of strategy for game theory is that it defines what decision a player will make given the situation they are in. This is known as a pure strategy concept. There can be probabilities assigned to when a player will use each pure strategy known as a mixed strategy concept.

In terms of mix zones, strategies will play a part in defining what decisions vehicles in the system can make. A vehicle's chosen strategy will affect how other vehicles in the systems decide to choose their own strategy.

2.4.3 Payoffs

Payoffs in game theory are known as a numerical value that a player obtains depending on what decisions were made by all players in the game. The model of payoffs is usually seen as a value that players wish to maximize as it represents the profit gained from making a particular decision. These payoffs coincide with a player's strategy, as a player will make a decision based on how they can maximize their payout in a game.

In the VANETs model and with the goal for security, the payout that vehicle nodes in the network will want to maximize the payoff of their own individual location

privacy. This payoff is calculated by what is gained following an exchange inside a mix zone. When a new type of adversary vehicle is introduced into the system, a new payoff will have to be defined for that type of attacker.

2.4.4 Nash Equilibrium

Nash equilibrium in a game represents a state of a game where players can no longer gain anything by changing their strategy based on the equilibrium of other players in the system. Players in the game have reached a Nash equilibrium if no single player can increase their payoff if they were given the advantage of knowing the strategies of all other players in advance. The Nash equilibrium consists of the optimal strategies for each player and the payoffs received for those optimal strategies.

In mix zones, vehicles in the system will gravitate towards a state of equilibrium. Vehicles will make the optimal decisions based on other vehicle's portrayal of optimal decisions in order to maximize the payoff they receive. Eventually, all the vehicles will be in a scenario where they cannot optimize their goal of maximizing location privacy any further. Later in the thesis when the attacking vehicles are introduced, the concept of equilibrium will be analyzed between the actions of all of the vehicles in the system.

Chapter 3

RELATED WORKS

Pseudonym switching and mix zones have been a recently developing concept within the field of VANETs with significant research going into both the security of the actual method of performing pseudonym switching and analysis of the effects of mix zones. Past works laid a foundation where problems and solutions exist for the topics were identified while shining the light as to what future research is necessary to advance security in VANETs. The following research has been outlined with important aspects towards the development of the content of my research.

3.1 Survey on Pseudonyms Changing Strategies for VANETs

This survey [3] evaluates and compares the different pseudonym changing approaches that exist for VANETs. Along with this, it also develops a discussion on the problems and challenges that exist within pseudonym changing strategies. The survey is one of the first to analyze this topic and start to push towards what is necessary for future research.

Some of the important topics analyzed throughout the survey that are particularly relevant to the contents of this thesis are:

- Metrics used to evaluate privacy
- The concept of an adversary model
- Costs involved in changing a pseudonym
- The existence of non cooperative behavior

The survey provides an introduction to the current research in these topics and lays the foundation for the advancement of future research.

In order to have a measure of effectiveness of pseudonym switching, there needs to exist a way to quantify and qualify the privacy in the system. Some of the most used metrics mentioned in the survey are anonymity set size, the entropy of the anonymity set size, the adversary's success rate, the maximum tracking time, and statistics on pseudonym changes. Incorporating these types of data allow one to be able to start identifying when the interaction of pseudonym switching is working given a particular model and when the expected level of privacy is not being maintained.

The survey also describes the use of an adversary model in a VANET system, where attacks may exist in order to try diminishing privacy in the system. It distinguishes different types of adversaries ranging from being global or local, active or passive, and internal or external. One of the main types of adversaries described is the location privacy adversary which tracks vehicle nodes by eavesdropping on the communications regarding those particular nodes.

The demonstration that there are costs involved when dealing with changing a pseudonym provides a concept that pseudonym changing is not a method providing full benefit. Some of the costs of conducting a pseudonym change are impact on the road safety depending on the type of strategy used, overhead costs necessary to carry out the pseudonym change, and possible loss of accountability of using a particular strategy.

A non cooperative behavior is developed from the costs involved in a particular strategy. Because cooperation between vehicle nodes plays a large role on how successful a pseudonym changing strategy is, it is important to recognize that there may not always be complete cooperation between all the nodes at a given time. This gives an introduction into using game theory as a means to calculate the gained benefit

versus the costs for a particular node given a particular scenario in order to arrive at a node's decision of whether or not to cooperate.

3.2 Analyzing Attacks and Defences for Security in VANETs

This paper [10] focuses at creating a comprehensive security analysis for VANETs due to the increase in security and privacy problems occurring since the development of VANETs. One of the key aspects explains that previous studies fail to examine the attackers' and defenders' costs and gains when modeling which leads to misinterpretation of the interaction between the attackers and defenders. The authors give a overall overview of different types of attacks and defenses that currently exist in the VANET environment

One of the important considerations that this paper makes is introducing the concepts of Return on Investment (ROI) and Return on Attack (ROA). The Return on Investment for the defense comprises annual expected loss, risk mitigation and cost of investment. The Return on Attack comprises the expected gain on attack and cost of the attack used to calculate how much an attacker could expect to gain from acting on a specific attack. The Return on Attack becomes an important attribute in how future research on adversaries on the privacy of VANETs is analyzed in order to determine when an attacker could choose to carry out an attack in order to maximize the potential privacy loss in the system.

3.3 Understanding Non-Cooperative Behavior using Game Theoretic Models for Location Privacy

One of the important aspects developed throughout this thesis is the concept that vehicle nodes can be modeled with non-cooperative behavior rather than with

fully cooperative behavior. In this paper [11], the authors describe that in terms of changing pseudonyms in mix zones, vehicle nodes act with non-cooperative behavior. Unlike previous models examining the effects on location privacy under the assumption that nodes that enter a mix zone will always cooperate with each other by switching pseudonyms, the paper develops the concept of a non-cooperative model for the nodes.

The non-cooperative properties for the nodes result from the properties that realize that changing pseudonyms in a mix zone can cost a significant amount of resources for a node. Therefore, due to such cost, nodes develop selfish behavior where they can choose not to cooperate in switching pseudonyms if the cost of changing is greater than the location privacy that would be gained for that node. This develops into the need for a way to model the result of how vehicle nodes would interact with each other given that all nodes are acting selfishly.

The paper applies game theory concepts in order to model the actions of selfish nodes, where the goal of each individual node is to maximize its own location privacy while minimizing the cost. One of the main developments used for the game theoretic model is a model of user-centric location privacy implemented to determine a node's location privacy over time. Using this model along with a location privacy loss function presented in the paper, the authors develop the payoff for a selfish node to be used the game theory model. Incorporating the concepts of equilibrium in game theory, the authors develop the optimal strategies for a selfish node for different types of games where a node could not gain any more location privacy from changing strategies. The authors examine this over 2-player games and n-player games along with complete and incomplete information games and present models for each type of game.

The paper leaves room for expanding on the model presented especially when

looking from a security standpoint. As the type of player presented in the non-cooperative model is that of a selfish vehicle node, taking a security perspective presents opportunities for introducing new types of players to the model which can more realistically replicate other types of players that may exist in real-life scenarios of VANETs. These new players could drastically change the resulting strategies presented in the games themselves.

3.4 Adversarial Presence in Mix Zones through Eavesdropping

This paper [13] investigates how mobile nodes can use mix zones to protect their location privacy in relation to when faced with an adversary. The aim of an adversary in the presented model is to track the location of the nodes overtime which in turn decreases the location privacy of the collective nodes. The authors use a game theoretic model to describe the interactions between the adversary and defending nodes with different strategies for each type of the different players. In order to accomplish the goal of the adversary, the adversary can use eavesdropping stations in order to attempt to track the nodes.

In terms of game theoretic concepts, the authors examine both complete and incomplete information games in terms of a small scale and large scale model of interactions. Analyzing the games at the small scale of one individual intersection in the network, the authors discover the single Bayesian Nash equilibrium at one intersection. Using this, they expand it to a large scale model of numerous intersections in a network before developing an algorithm to obtain the equilibrium over the larger network. This examination of the interactions on both a micro and macro scale is an important concept in examining how to expand the analysis of VANETs.

This paper successfully introduces the concept of having a malicious adversary in a particular technological environment of mix zones in VANETs. Often past research

only examine the use case of inherent users of the system which in this scenario are mobile nodes acting in the mix zones. The inclusion of an adversary trying to eliminate the goal of the inherent users of the system encompasses the concept of security in computer systems. It introduces a means to connect the attackers and defenders in a system and further moves into solving the problem on how researchers can prevent attackers from successfully given the interactions between the attackers and defenders. Along with this, the discovery of how the lack of knowledge of the mobile nodes of the attacker and the attacker's strategy leading to a decrease in location privacy shows the potential effects of an attacker's effect on decreasing location privacy in a system.

3.5 Modeling an Attacking Vehicle Inside the VANET

Though many types of attacks on the privacy of VANETs focus on obtaining tracking information from outside of the vehicle node system, this particular paper [19] works a out a model that describes what happens when an attacker node is included inside the collection of defender vehicles inside a VANET. This particular model describes the interactions of an N + 1 vehicle node system, where there exists N defender vehicles and one malicious vehicle in the system.

The difference demonstrated by using this type of model allows the attacking vehicle to effect the system on an internal level, where the attacker is a vehicle within the system. The security implications of this type of model show a demonstration of a new type of threat modeling for VANETs. It brings the question of how adequate are the defenses previously modeled work against a possibly unknown attacker inside the system. This pushes forward necessary future research towards how defending vehicles could attempt to detect an anomaly inside the system and whether the network has been compromised by an adversary.

Although the paper goes into more in-depth mathematical explanation of the game developed from the N + 1 node model, the basis of model from a security perspective contribute a novel point of view of how to protect the privacy of vehicles in VANETs.

Chapter 4

DESIGN

In this chapter, I describe the design of the model for the interaction of greedy vehicle nodes with the inclusion of internal attacker vehicle nodes. I focus on describing the overall goals and requirements of the model in order to accurately portray the results that would occur when applied to a real VANET system. In chapter 5, I explore a detailed implementation of this model.

4.1 Goals

The primary goal in creating the model is to demonstrate the effects of adding byzantine attackers to a system of greedy vehicle nodes [14]. This will be used to help build on the severity and impact that internal attackers can have on mix zones and allow future research to determine how to implement defenses to further secure mix zones from potential malicious adversaries.

4.2 Requirements

In this section, I define the necessary system requirements in order to accurately model what would be seen in actual interactions in a mix zone.

The following are the formal requirements that the mix zone model will maintain:

- My model must support scalability of the number of vehicles that can be in the network.
- My model must support dynamic changing of the amount of attacker vehicles currently in the system.

- My model must be able to quantitatively maintain the location privacy levels of the defending vehicles in the system

4.2.1 Scalability

The scalability of the number of vehicles in the network is critical in order to help model real life scenarios of the number of vehicles that exist in a VANET. When talking about roads and freeways, the amount of vehicles clustered together are in the hundreds and thousands with groups of vehicles all being in the same VANET. Because of this, large scaling of the number of vehicles and their interactions in the model will help mimic the potential size of VANETs.

4.2.2 Numerous Attackers

Another important requirement to consider in the model is that there is not only one single adversary trying to attack the system at a time. When talking about a standard computer network, there are often numerous attackers trying to infiltrate the network at the same time in order to maximize the severity of the effect of the attacks on the network. This observation can also transfer over on VANETs. In order to make sure that defenses can be created based on the potential of attackers, the model needs uphold the potential for multiple attackers that may exist in the system at a time.

4.2.3 Quantitative Location Privacy

In order to be able to determine what effect attackers have on the system, the model need to sufficiently track the decrease in location privacy of the system. The model must keep track of the current location privacy for each vehicle in the system, specifically the defending nodes in the system. By maintaining the location privacy

levels, the model will be able to summate the system's location privacy after exiting a mix zone and determine how much the attackers have affected the network.

4.3 Mix Zone Model

The modeling of the mix zone is the basis for evaluating the interactions between vehicles in the network. The following is the process for vehicles follow for one instance of entering and exiting a mix zone.

1. There are n vehicles that enter the mix zone.
2. After the vehicles enter, each vehicle makes a decision whether or not it wants to cooperate in switching pseudonyms or defect in switching pseudonyms.
3. The vehicles that do decide to cooperate switch pseudonyms along with all other cooperating vehicles.
4. The vehicles that defect take no action in the mix zone.
5. After this process, the vehicles exit the mix zone.

Figure 4.1 shows an overview of this process for a 20 node network (a) where 10 nodes entering a mix zone (b) and out of the those 10 nodes in the mix zone, 7 of the nodes decide to cooperate with each other and 3 of the nodes decide to defect (c).

There are a list of rules and assumptions that are associated with the model regarding mix zones. To summarize the process with the rules and assumptions stated above, the following algorithm is created for how a vehicle decides to cooperate or defect in a mix zone:

1. When a vehicle enters a mix zone, it observes the number of vehicles in the mix zone and its current location privacy

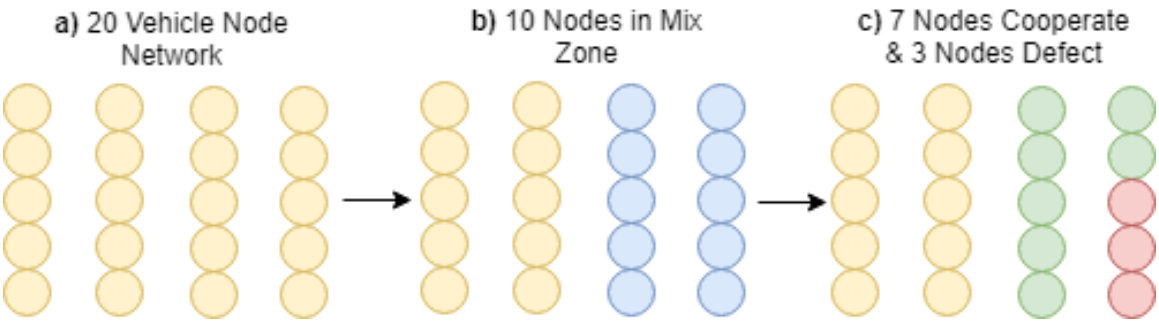

Figure 4.1: Model for Process of a Mix Zone for a 20 Vehicle Node Network

2. It formulates the potential gain in location privacy based on those values and a predictive model independently of other vehicles' decision.

3. It decides whether or not to cooperate in the mix zone based on the calculated value.

4. After the vehicle exits, it observes how many vehicles decided to cooperate to calculate deviation for suspicion.

4.4 Location Privacy Model for the Mix Zone

One of the key aspects for modeling the mix zone is how the location privacy gained is determined. In previous research [11], there has been a mathematical formula defined to quantitatively determine the location privacy that will be gained in a mix zone.

That formula is:

$$LP = \log_2 n \tag{4.1}$$

where

$$n \geq 2 \tag{4.2}$$

LP represents the location privacy of the cooperating vehicle nodes after the mix

zone and n represents the number of vehicle nodes that cooperated in the mix zone.

One of the key concepts that comes from the formula (4.1) is that there is an upper bound on the maximum amount of location privacy a node can have in the network. This is bound by the maximum value of n or the number of vehicle nodes in the network. The condition (4.2) represents that in order for pseudonym switching to take place in a mix zone, there must be at least two nodes in the mix zone since that is the minimum for nodes to be able to switch with each other.

An example of this formula with a 50-vehicle node network is shown in Figure 4.2. The upper bound of location privacy for a node in a 50-node system is approximately 5.64. As a result, the location privacy of any node in the network at any given time is between 0 and 5.64.

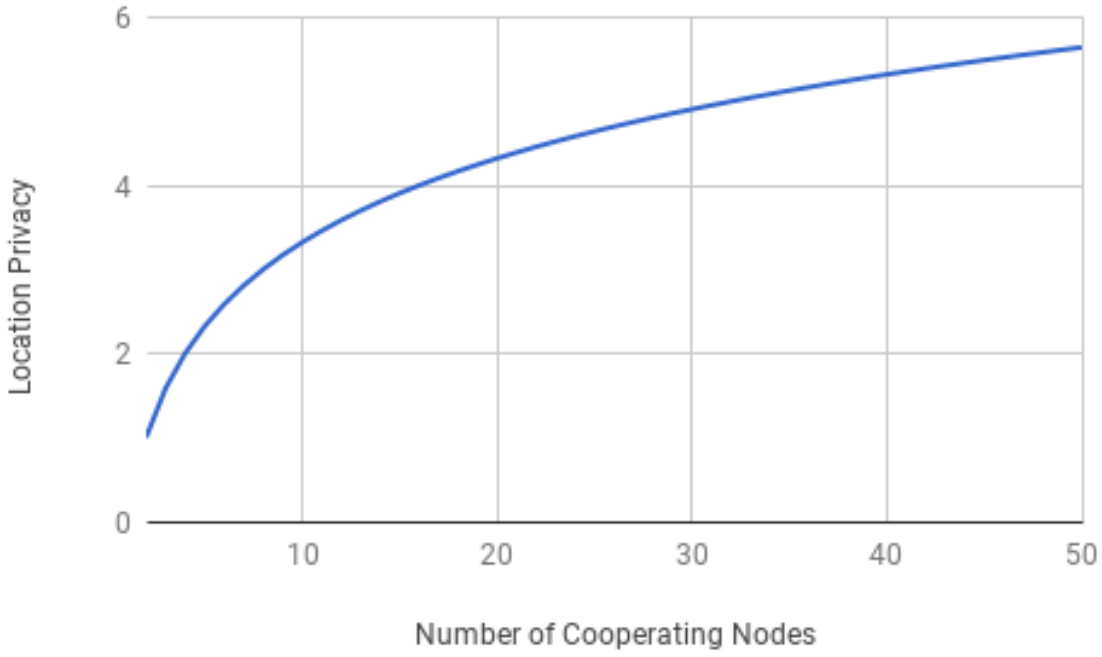

Figure 4.2: Location Privacy of a Node After Mix Zone in a 50 Vehicle Node Network

4.5 Location Privacy Loss Function

Another important aspect of the design of the model is the property of location privacy being lost over time. When nodes do not switch their pseudonyms, they become more susceptible to being location tracked. This is because their identifier remains constant and thus becomes easier for outside adversary to connect a vehicle to its broadcasts. Because of this, the model needs some way of modeling a loss in location privacy as time goes by where a vehicle doesn't cooperate in pseudonym switching. In this model, the loss in location privacy is modeled as a constant linear decrease. This means that for each instance of a mix zone, regardless of whether or not a node in the network enter the mix zone, cooperate or defects, there is a constant loss in location privacy. Since the location privacy of a node must be at least zero, if a node never switches pseudonyms, the minimum location privacy it will have will be at least zero. The model incorporates this function to force nodes to eventually switch pseudonyms if they want to maintain a high level of location privacy.

The location privacy loss function is:

$$LPLT = x * t \tag{4.3}$$

where $LPLT$ is the location privacy loss over time, x is a constant for location privacy loss and t represents the time that has past or in this particular case, the number of rounds that have past. From this, the formula (4.3) summarizes that between each round of a mix zone, there is a constant loss in location privacy for a vehicle.

4.6 Vehicle Types for the Mix Zone

In our model, there exist two different types of vehicles that can be in the network for mix zones. The first type is the defender node. These nodes represent the vehicles

that are trying to maintain their own location privacy in the network in order to maintain security. The second is the attacker nodes which represents the new type of vehicle that did not exist in the previous models. They are trying to lower the location privacy of the overall system and harm the security of the defender nodes. These vehicle types represent the concept of different types of players in our model of mix zones. The following parts provide an in-depth analysis of each type of vehicle in the model.

4.6.1 Defender Vehicles

The defender vehicles represent normal user vehicles in the network. These vehicles represent the maintenance of location privacy in the network. Ultimately, attackers are trying to affect the location privacy of the defender nodes in order to achieve their goal.

The following represent the overall goal of an individual defender node in the model:

- A defender vehicle wants to maximize its own individual location privacy in the network
- For a defender vehicle, the location privacy of other defender vehicles in the network or the location privacy of the network as a whole is insignificant.

The concept of the defender node describes that the nodes act selfishly in a mix zone. If there is no potential gain in location privacy for a defender for cooperating in a mix zone, then a selfish defender will not cooperate. Because all defender vehicles in the model are all selfish, then no defender cares about the overall location privacy of the network.

The following assumptions is also made in terms of the knowledge of defender vehicles:

- The defender assumes that all vehicles in the network are also acting as defender nodes.
- A defender maintains a certain level of suspicion based on the deviation of the predicted cooperation probability.
- A defender has no knowledge of the existence of any attacker vehicles in the network unless raised by suspicion.

These assumptions are necessary into creating a model relevant to how adversaries act in computerized systems. Often, attackers will make sure they remain undetected in a system. In order to start with the basis of the model, the assumption for defenders to assume that the network only consists of greedy vehicle nodes is used.

4.6.2 Attacker Vehicles

The attacker vehicles represent the adversarial vehicles in the network. These vehicles represent the possibility of malicious actions to tamper with the location privacy of a VANET.

The following represent the overall goal of an individual attack node in the model:

- An attacker vehicle wants to maximize the location privacy loss of the defender vehicles in the network.
- An attacker vehicle wants to avoid being detected by the defender vehicles.

This goal of the attacker vehicles in the model represents what malicious adversaries work for when dealing with computerized systems. Since the defender nodes want to

maintain their own location privacy in order to sustain their security, attackers wish to threaten the security of the network by lowering the overall location privacy of the network. By lowering the location privacy, the network becomes more susceptible to location tracking by other types of attacks on the VANET. This leads to the benefit of having this type of internal attacker vehicle within the network. An internal attacker vehicle that can lower location privacy benefits chaining that attack to other outside attackers on the network.

4.7 Strategies

As mentioned earlier, there exist two different strategies that vehicles can decide on inside the mix zone. Each vehicle can decide between two different strategies:

- Cooperate: Vehicle decides to switch pseudonyms
- Defect: Vehicle decides not to switch pseudonyms

These strategies apply to both defender and attacker vehicles. However, the benefit of each strategy is dependent on the type of vehicle. The next section describe the reasoning why a vehicle would decide a certain strategy given a particular situation in a mix zone.

4.7.1 Defender Strategies

A defender would choose to cooperate in the case that it would be able to increase its current location privacy through cooperating. However, the location privacy gained after cooperating in a mix zone depends on the number of cooperating nodes and that a defender node does not know how many nodes will cooperate before choosing a strategy. Because of this, a defender node has to predict how many

nodes in the mix zone will decide to cooperate. If the defender predicts that it would gain location privacy through cooperating based on a predicted amount of cooperated nodes, then the defender will decide to cooperate.

A defender would in turn decide to defect in the case when it could not gain any location privacy through cooperating. When a defender decides to defect, it only cares about the value the decision is bringing to itself. However, the decision it makes affects the amount of location privacy gained from the other nodes that decide to cooperate. Since a defender chooses a strategy based on a predicted number of cooperating nodes, if the actual number of cooperating nodes is lower than the expected number of cooperating nodes, then the cooperating nodes might gain less location privacy than expected.

Since all defender nodes are following the same goal of maximizing their own location privacy, any deterrence to that goal is important to the defender. The defender nodes all follow the same model of prediction cooperating nodes. This means that if the VANET truly exists of only selfish defender vehicle nodes, then there should be little deviation from the prediction model. This property will be expanded in a later section.

The equation for a defender's strategy is a mixed strategy equation:

$$S_{def_{x,y}} = (p, 1-p) \tag{4.4}$$

where x is the strategy where the defender cooperates which is played with probability p and y is the strategy where the defender defects which is played with probability $1-p$.

4.7.2 Attacker Strategies

An attacker would choose to cooperate in the case when it wants to avoid being detected by other defender nodes. Even though cooperating means that the amount of location privacy gained for the cooperating defender nodes would increase, the priority of not being detected may cause an attacker to decide to cooperate. Another cause of decision for an attacker to cooperate is that the number of nodes entering the mix zone is low. Since an attacker wants to maximize the location privacy loss in a system, it wants to affect the highest of defender nodes when it chooses a strategy. Therefore, cooperating when a low amount of nodes are in a mix zone will provide a lower benefit for defender nodes than cooperating when a high amount of nodes are in a mix zone.

An attacker would choose to defect in order to lower the location privacy gained by cooperating defender nodes. As explained by the location privacy model from cooperation in a mix zone, the greater the number of cooperating nodes is, the greater the location privacy gained from the mix zone is for each cooperating defender. Therefore, when an attacker decides to defect in a mix zone, it is taking away potential location privacy gained for cooperating defender nodes by a potential of one additional node. However, because defender nodes are predicting a certain amount of cooperation, any deviation of an attacker's choice in strategy from that of a selfish defender node can raise suspicion that there is an anomaly in the network.

Attackers have to mix the choice of strategy in a mix zone in order maximize the privacy loss while remaining undetected. This mix in strategy for the attacker is what the model will help determine. The equation for an attacker's strategy is a mixed strategy equation:

$$S_{att_{x,y}} = (q, 1 - q) \tag{4.5}$$

where x is the strategy where the defender cooperates which is played with probability q and y is the strategy where the defender defects which is played with probability $1 - q$.

4.8 Payoffs of Vehicles

With all information about the goals of each type of vehicles and modeling how location privacy is obtained throughout the system, the model for payoffs takes into consideration these values. Since the defender and the attacker nodes have different overall goals, the payoffs for each consider different attributes of the model.

4.8.1 Defender Payoff

When talking about the payoff of a selfish defender vehicle node in the network, the model will examine if there is an overall benefit to the defender in increasing its location privacy from its current state. There are three main aspects to consider which will determine a defender's payoff from deciding to cooperate in a mix zone. They are:

- The location privacy gained from cooperating with the actual number of cooperating nodes.
- The current location privacy of the defender node.
- The cost it takes to change a pseudonym.

There exist certain costs when changing pseudonyms. This includes the cost of acquiring new pseudonyms, the cost of updating the routing tables of the middleware during the switch, and the cost of remaining silent while inside a mix zone [11]. They combine into a singular cost which can be demonstrated as a loss in location privacy.

These aspects will be used to decide whether or not a defender should cooperate or defect.

Another consideration in this payoff model is that the defenders do not know in advance how many actual nodes will cooperate in the mix zone. Since they can only estimate the number of nodes that will cooperate, this brings a new issue where even if a defender node chooses a strategy based on an expected payoff, they may receive a payoff that was less or greater than expected. This uncertainty comes into play when talking about the level of suspicion defenders may confront.

For the defender payoff, it is defined as follows:

$$ELPG = ELP_{after} - LP_{before} - P_c \tag{4.6}$$

where $ELPG$ represents the expected location privacy gained for a node, ELP_{after} represents the expected location privacy of the node after the mix zone assuming cooperation, LP_{before} represents the location privacy of the node before the mix zone, P_c represents the cost of a pseudonym change.

The evaluation of the formula (4.6) determines whether defender vehicle will decide to cooperate or defect for the defender's strategy. The rules for determining the evaluation are as follows:

- If $ELPG > 0$, then the defender vehicle chooses to the strategy to cooperate
- If $ELPG < 0$, then the defender vehicle chooses to the strategy to defect

4.8.2 Attacker Payoff

To describe the payoff of an attacker, the consideration takes into account the overall goal of the attacker to maximize the amount of privacy loss gained by the defender vehicle nodes in the system. In order to account for this, the payoff is

a difference between location privacy when the attacker vehicle cooperated versus when the attacker vehicle defected. This difference is summed over all the cooperating defender nodes that have been affected by the attacker's strategy. For an attacker that defects, this value will provide a positive payoff for the attacker. If the attacker cooperates, the model will receive a zero payoff based on not causing any location privacy loss.

The question that forms out of this payoff model is why would an attacker ever cooperate if they would always receive a positive payoff. This is where the attacker's second goal of not being detected comes into the payoff model. Being detected in the network is the downfall for the attacker and that's why the attacker can not always defect as to not raise suspicion.

For the attacker payoff, it is defined as follows:

$$TLPL = (n - a) * (LP_n - LP_{n-a}) - D \tag{4.7}$$

where $TLPL$ represents the total location privacy lost for all cooperating defenders in the mix zone, n represents the number of nodes (all defenders and attackers) in the mix zone, a represents the number of attacker nodes that defect in the mix zone, LP_n and LP_{n-a} represents the location privacy that would have been gained if there are n and $n - a$ cooperating defender nodes respectively, and D is a value representing the level of possible detection the adversary is currently facing. The value of the level of suspicion of defenders for detection is described in a later section.

4.9 Expected Probability of Cooperation Function

As mentioned above, the defender nodes have an expected number of nodes that cooperate that they use in order to predict the expected payoff. For this, the model needs a way to model the expected probability of cooperation with a function. The

requirements for the function should match that of a probabilistic function. These are the requirements that the function has:

- The function should output a value between 0 and 1 inclusively.
- The function takes into account that nodes are more likely to cooperate if the number of nodes in the mix zone is close to the total number of nodes in the VANET.

Taking these requirements in the function, the expected probability of cooperation is modeled as followed:

$$EPC = \log_2 n / \log_2 m \tag{4.8}$$

where

$$n \geq 1 \tag{4.9}$$

and

$$m \geq n \tag{4.10}$$

EPC represents the expected probability of cooperation of the nodes in the system, n represents the number of vehicle nodes that are in the mix zone, and m represents the total number of vehicle nodes in the VANET. The condition (4.9) needs to be satisfied as there needs be at least one node to predict cooperation. The condition (4.10) shows that the amount of nodes that join the mix zone cannot exceed the number of nodes in the VANET.

An example of the expected probability of cooperation function with a 50-vehicle node network is shown in Figure 4.3. As seen, the higher number of nodes that enter a mix zone results in a very high probability for each node to cooperate with all 50 nodes in the mix zone representing a 100 percent expected cooperation rate where when 2 nodes joining a mix zone out of the 50 total nodes, only a 17 percent cooperation rate is expected.

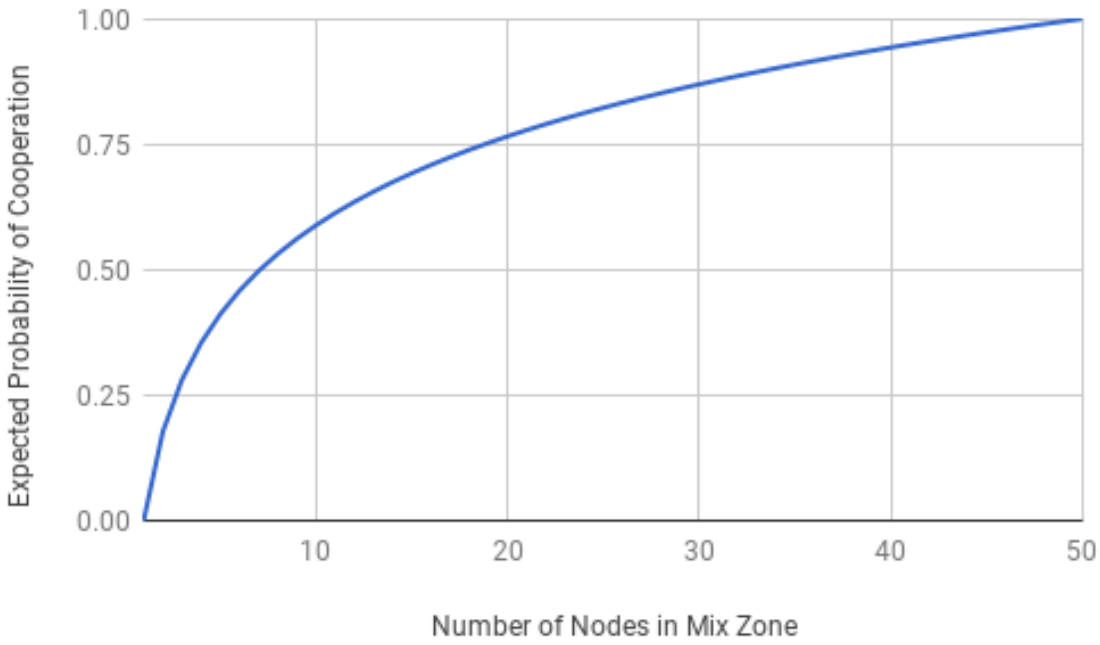

Figure 4.3: Expected Probability of Cooperation of Nodes in a 50 Vehicle Node Network

4.10 Level of Suspicion

The last aspect of defenders that the model needs to consider is how the defenders will attempt to detect if there is an anomaly in the network. The way the model considers this is by detecting if there is a larger than normal deviation from the expected probability of cooperation. If an attacker were to defect an abnormal amount more than that of a selfish defender node, then the deviation would become larger. At some point, defender nodes would consider a deviation over a certain amount to be suspicious and thus detect attackers in the system. Because deviation data from one instance of a mix zone is not accurate, the average deviation over multiple iterations of mix zone instances can be taken. The value at which defenders detect attackers depends on how suspicious the defenders are in the system.

4.11 System Summary with Formal Game Model

The formal game for the design is defined as G which is defined as a triplet (P, T, S, U) [11], where P is the set of players, T is the set of types of players, S is the set of strategies, and U is the set of payoffs. P consists of n vehicles in the network, where each vehicle in n is either defined as an element in T: either a defender or an attacker. S consists of two different strategies that each player can choose from: cooperate or defect.

The following figure 4.4 shows a formal game model for the payouts of both types of players for a two player game. However, the difference in the figure from the formal game model is that the figure only shows the interactions of two players of one defender and one attacker. In the formal model, there are n players where each player's strategy in the game determines the payoff for all the other players. Given equations 4.3, 4.6 and 4.7, I develop a basis for a formal game model for the attack and defense scenario for the interactions that occur in a mix zone.

		Defender: Cooperate	Defender: Defect
Attacker	Cooperate	[ELP_{after} + LP_{before} - P_c -LPLT, 0]	[-LPLT, 0]
	Defect	[ELP_{after}+ LP_{before}- P_c , (n - a) * (LP_n - LP_{n-a}) - D]	[-LPLT , (n - a) * (LP_n- LP_{n-a}) - D]

Figure 4.4: Formal Game Model for Defenders and Attackers. 2-player Game with Cooperate and Defect Strategies for Both Attackers and Defenders and Their Corresponding Payoffs

Chapter 5

IMPLEMENTATION AND RESULTS

This chapter explains the implementation and results for the design of the model that was described previously. The explanation goes through the process of multiple models that leads up to the final model of the interaction between the success of attackers on maximizing location privacy loss while remaining undetected by different levels of defender's suspicion. As well, the final model evaluates a solution for the optimal strategy for attackers. All the simulations for the models were implemented in Python.

5.1 Variables

There are numerous variables that the model can control to simulate various scenarios of VANETs. Changing these variables can lead to different effects on strategies each type of vehicle nodes chooses. Variables are tested in the models to analyze possible trends of the results. The variables tested are described below.

5.1.1 Total Number of Nodes in the VANET

The total number of nodes in the VANET can be adjusted to simulate that of a real scenario VANET. In order to obtain accurate results with less variability, the model uses a large number of nodes in the VANET. If the model uses a small number, the results would contain less data that could be analyzed to determine the effects of attackers in the system. This number also determines the maximum possible location privacy for a single node based on formula 4.1.

5.1.2 Cost of Pseudonym Change

The cost of a pseudonym change occurs anytime a defender node decides to cooperate in a mix zone. The higher the cost of a pseudonym change, the higher the chance that a defender node will not cooperate in the system due to the negative effect of the payout. In the models, the cost of pseudonym change is synonymous with γ.

5.1.3 Number of Rounds of Mix Zones

The number of rounds of mix zones determines how many instances of mix zones occur in the simulation. One round represents an instance where a certain number of vehicle nodes enter a mix zone then exit a mix zone. After every round in a mix zone, the location privacy for each defender node is reevaluated. A larger set of rounds provide a better idea of the average interaction between the defenders and the attackers.

5.1.4 Chance to Enter Mix Zone

The chance to enter the mix zone determines the probability of each individual node in the VANET to enter the mix zone in a given round. If the chance to enter the mix zone is 10 percent, then in each round, every vehicle node has a 10 percent chance to enter the mix zone. In a 50-node network, an average of 5 nodes will enter the mix zone. However, when analyzing each round individually, the amount of nodes that enter the mix zone will significantly vary.

5.1.5 Location Privacy Loss Per Round

The location privacy loss per round represents the location privacy loss function mentioned in the design phase of the model. Between each round of mix zone in the simulation, the location privacy of each node in the VANET will decrease linearly by the specified value. If the location privacy of a node were to decrease below zero after a round, then the location privacy of that node will remain zero.

5.1.6 Probability Error Adjustment

The probability error adjustment represents a manual adjustment to the expected probability of cooperation function in order to closer match the actual probability of cooperation adjustment in the simulations. This value depends on all other variables and is calculated and adjusted over simulations. It allows the average difference of the expected probability of cooperation and the actual probability of cooperation to be zero.

5.1.7 Minimum Number of Nodes that Enter a Mix Zone

The minimum number of nodes entering a mix zone represents the specified amount of nodes for a round of a mix zone to be valid. This number is used in order to mitigate the skewed data that comes from the logistic functions utilized for small amount of nodes.

5.2 Initial Selfish Defense Node Model

In order to start creating a model inclusive of attackers, the first implementation is to create the initial model showing how defender nodes interact with each other based on the selfish nature of vehicles for mix zones. The goal of this model is to

determine whether or not a defender vehicle will cooperate or defect depending on the current state of certain variables.

The dynamic variables for this model include:

- The total number of nodes in the VANET
- The number of nodes that enter the mix zone
- The current location privacy of the defender node
- The cost to switch pseudonyms

Based on these variables, the model calculates the defender payoff is in formula 5.1 based on the expected probability of cooperation. If that value is positive, then the model will output the fact that the defender node would cooperate in the mix zone. If the value is negative, then the model will output the fact the defender node would defect in the mix zone.

Figure 5.1 demonstrates the interaction among all of these variables to produce a chart showing when a defender would choose to cooperate versus defect as the defender's strategy. If the current location privacy of the defender node is in the green area based on the y axis, then the defender would choose the cooperative strategy. If the current location privacy of the defender node is in the red area based on the y axis, then the defender would choose the defecting strategy. If the max location privacy possible is approximately 5.64, with the cost of the pseudonym change being 0.3, the maximum possible location privacy that could be obtained from cooperating would be 5.34. In an observation, there are certain amount of nodes, specifically low amounts, in a mix zone where a defender node would never cooperate because the location privacy gained after the mix zone would be zero.

Figure 5.2 demonstrates a similar concept but with a cost of pseudonym change

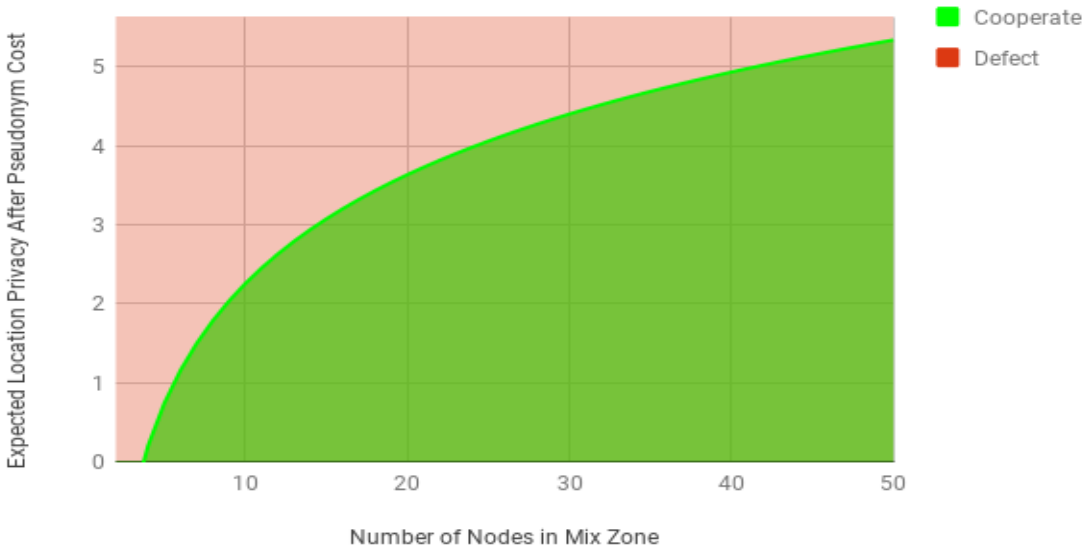

Figure 5.1: Graph of Defender Strategy in a 50 Vehicle Node Network with $\gamma = 0.3$

value of 1.0. Because the cost of changing a pseudonym is higher than the previous chart, there is a larger amount of area where a defender node would defect because of less benefit from a mix zone

5.3 Initial Selfish Defense Node Model With Attacker Nodes

Since the basic self defense node model has been established, the next step in implementation is to introduce an attacker nodes into the system. The goal of the model is to determine whether or not an attacker would cooperate or defect , giving the variables defined from the previous model. Based on these variables, the attacker obtains the payoff of the difference in expected location privacy for all defender nodes between if the attacker would have cooperated and if the attacker would have defected. If this payoff is positive, then the attacker would defect. If this payoff is negative, then the attack would cooperate.

In Figure 5.3, the expected total location privacy loss when a single attacker was

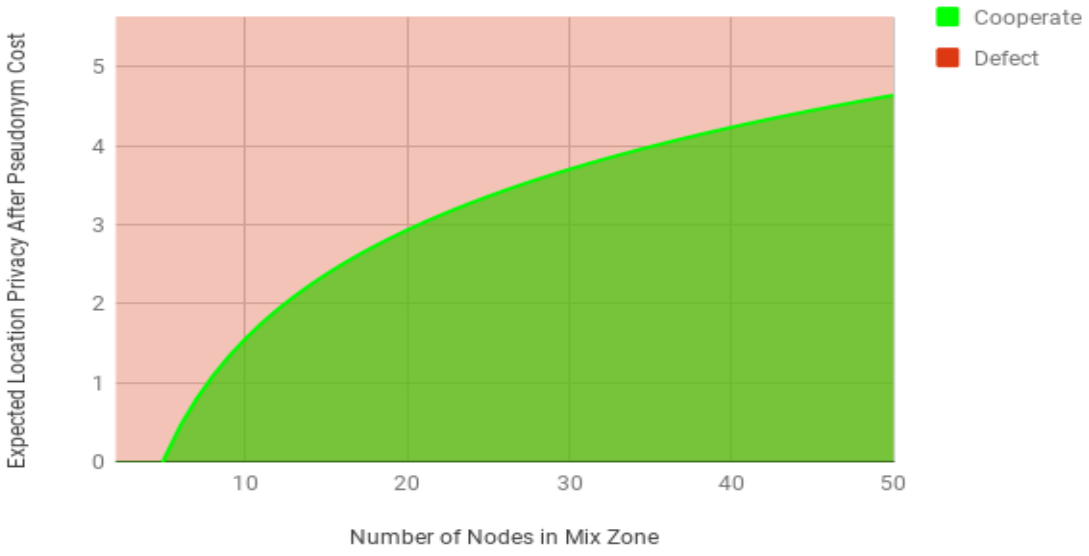

Figure 5.2: Graph of Defender Strategy in a 50 Vehicle Node Network with $\gamma = 1.0$

included on the selfish defense model is described. The displayed graph shows that the expected total location privacy loss is positive across all quantity of nodes that enter the mix zone. It means that for all instances of number of nodes that enter the mix zone, the attacker has no reason to choose any other strategy than to defect. In this particular graph, the maximum value of total privacy loss occurs with a low amount of nodes entering the mix zone.

In Figure 5.4, the number of attackers increases from one to three. What can be observed is that the total location privacy loss increases than that of a single attacker assuming that all three attackers decide to defect. Similar to the previous graph of the single attacker, the total location privacy loss remains positive for all iterations of the number of nodes that enter the mix zone. However, there is a slight difference in the peak of the graph. The maximum of total location privacy loss occurs at 16 nodes entering the mix zone rather than at the beginning numeral values of nodes.

The main take-away from these two charts is that an attacker has no reason to ever not defect as the strategy. As a result, it leads into understanding a reasoning

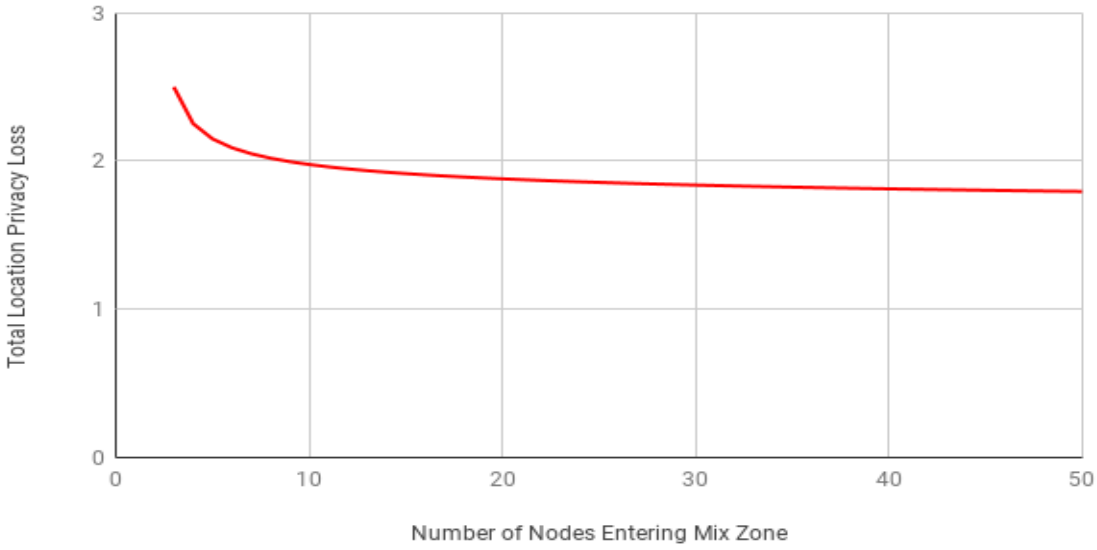

Figure 5.3: Chart of Expected Total Location Privacy Loss in a 50 Vehicle Node Network with 1 Attacker with $\gamma = 0.3$

for why an attacker would cooperate. An attacker may cooperate if there is a risk of being detected. This finding will be evaluated further in the implementations of simulations.

5.4 Selfish Defense Node Simulations without Probability Error Correction

With modeling how selfish defender nodes should cooperate and defect under specific conditions, the next goal in implementation is to run simulations of the game scenarios with the design model made previously. In addition to the variables used in the original model, there are some more variables needed in order to more accurately simulate the scenario closer to a real life VANET scenario.

The following are the dynamic variables used for the simulation:

- The total number of nodes in the VANET

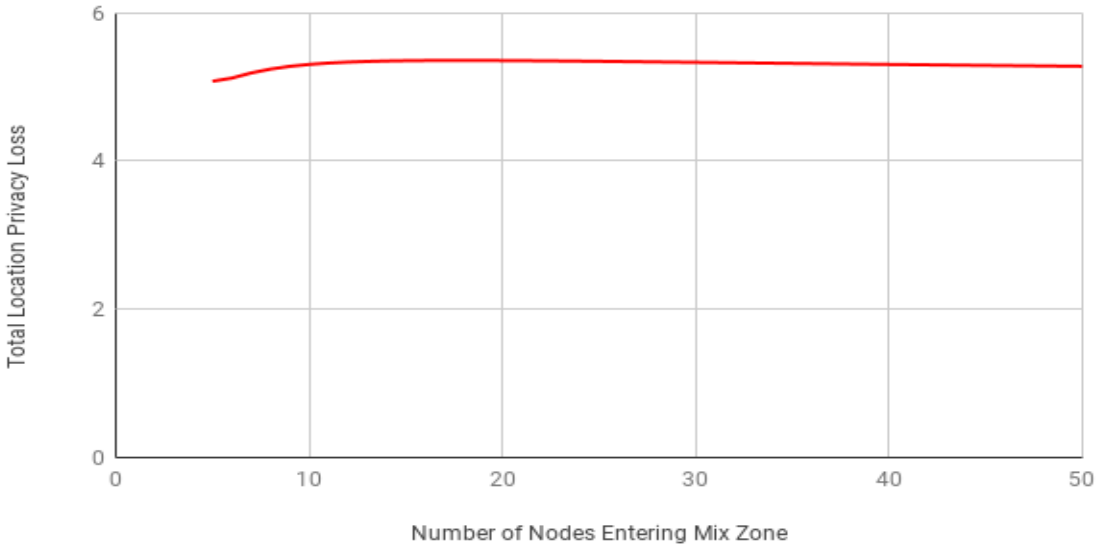

Figure 5.4: Chart of Expected Total Location Privacy Loss in a 50 Vehicle Node Network with 3 Attackers with $\gamma = 0.3$

- The minimum number of nodes that enters a mix zone
- The number of rounds of mix zone
- The cost to switch pseudonyms
- The chance to enter the mix zone
- The location privacy loss per round

The main goal of the first simulation is to evaluate and confirm that the expected probability of cooperation matches closely with the actual probability of cooperation which is simulated. In the initial simulation, there are only selfish defender nodes in the network. Each node starts with a random location privacy between zero and the maximum location privacy allowed based on the number of nodes in the network. For each round of the mix zones, each node in the network has whatever set chance to enter the mix zone. Based off the simulation of chance, there are n nodes that enter the mix zone per round. Each node in the mix zone evaluates its current location

privacy. Based off the previous expected model, it determines whether to cooperate or defect in the mix zone. The payoffs for the cooperating defender nodes are then updated based on the actual number of nodes that decide to cooperate. Lastly, the location privacy of all the nodes is decreased by the location privacy loss per round. This process is repeated in the number of rounds specified by the simulation.

Figure 5.5 describes one simulation of all the selfish defense vehicles model. This simulation is done over 10,000 rounds of mix zones when $\gamma = 0.3$, the chance to enter the mix zone is 0.1, the constant location privacy loss per round is 0.1, and with 50 vehicle nodes in the network. If the expected probability of cooperation is accurate on the modeling of the actual probability of cooperation, then the difference in probability should approach zero or close to an average of zero per round. However, when we look at the results from the graph, there is a difference of 210 between the expected probability and the actual probability of cooperation, where the expected probability overestimates the actual probability. When considering the average difference per round, it comes to 0.021 which is equivalent to an average difference of 2.1 percent per round. It can be considered a significant amount for the large number of rounds in the simulation. This leads to the need to find a way to correct the difference to become closer to an average of zero.

In addition to this result, other interesting data the simulation provides is that the average location privacy of a single node at any given time is approximately 1.23 for this particular simulation. Over multiple runs of the simulation, the average location privacy remains around that value with a deviation of about 0.01. An average of 22 percent of the maximum potential location privacy is held by a node at a time. This is considered a relatively low amount.

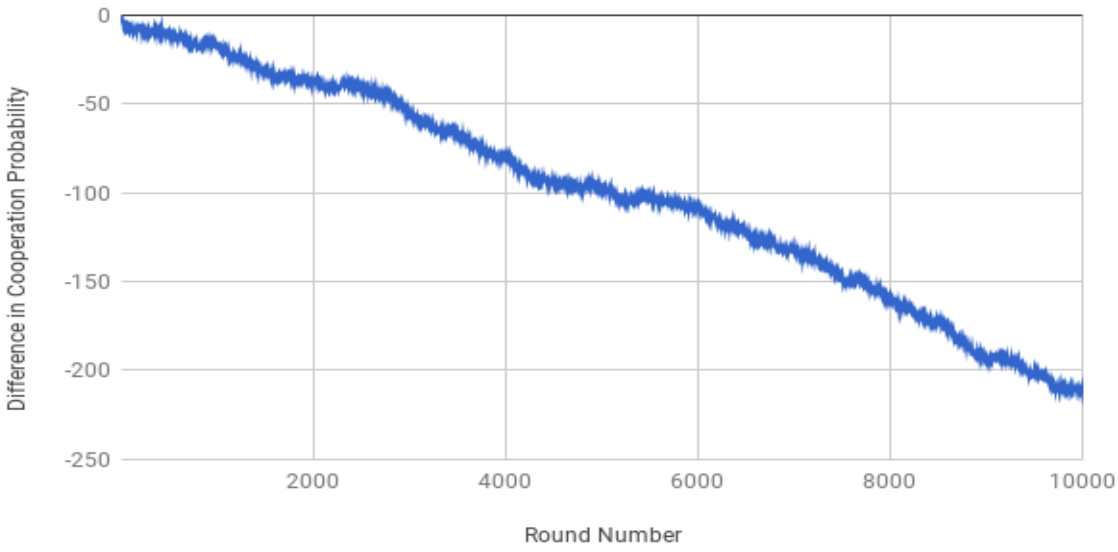

Figure 5.5: Graph of Difference of Expected and Actual Probability of Cooperation without Probability Error Adjustment in a 50 Vehicle Node Network with No Attackers

5.5 Selfish Defense Node Simulations with Probability Error Correction

An important aspect to examine is that defender nodes do not have any knowledge of the current location privacy level of other nodes. However, that location privacy level is personally taken into account when conducting whether or not to cooperate in a mix zone. Since local location privacy is taken into account in the strategy but not in the predictive model, the predictive model overestimates the actual cooperation rate. To solve this issue, a new model is created to adjust the expected cooperation probability by utilizing a constant amount to leverage the expected cooperation probability towards the actual cooperation probability. This probability error is formed from multiple simulations and from adjusting the error each time by the average probability difference per round which is formed over multiple iterations. This process is done until the average probability difference per round is sufficiently close to zero.

Figure 5.6 describes the new simulation with the additional probability error adjust. After adding the adjustment, the difference in probability only varies between -15 and 10 over the 10,000 rounds. This equates to a 0.0015 average deviation from zero per round. In comparison to the previous model, the average location privacy of a single node at any given time is approximately 1.15. Note that since the error correction is calculated per simulation with specific variables, changing around variables will cause discrepancy. As a result, the error correction has to be recalculated. When examining the average of this over 30 different runs with probability error adjustment, the average also is considerably closer to zero as shown in Figure 5.7.

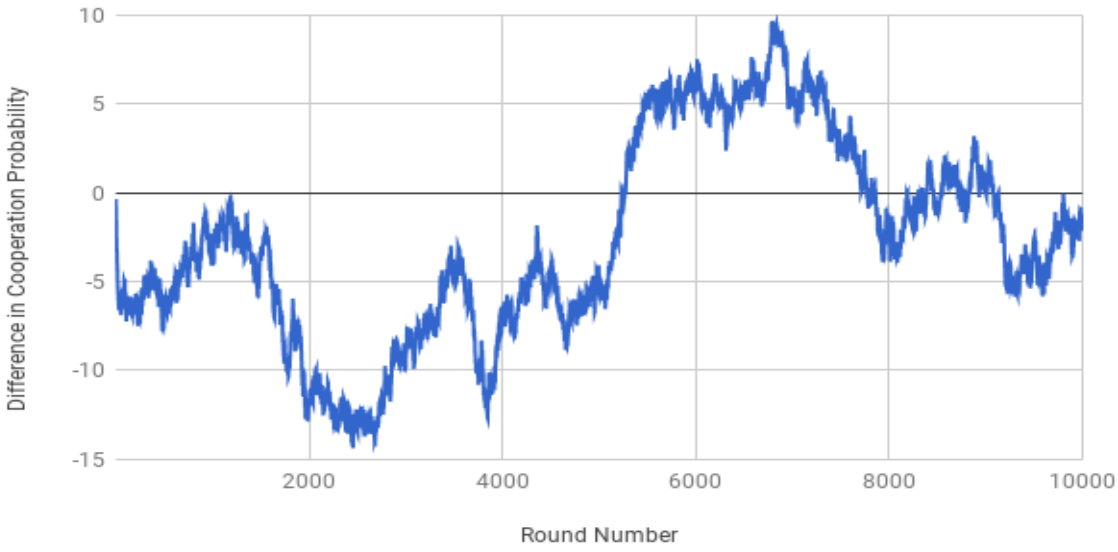

Figure 5.6: Graph of Difference of Expected and Actual Probability of Cooperation with Probability Error Adjustment in a 50 Vehicle Node Network with No Attackers

5.6 Adding Attackers to the Simulation that Only Defect

The next step for simulation implementation consists of adding the attackers in the model and determining how they affect the system. First of all, we have to determine the effect that the attackers have on the difference in cooperation probability when

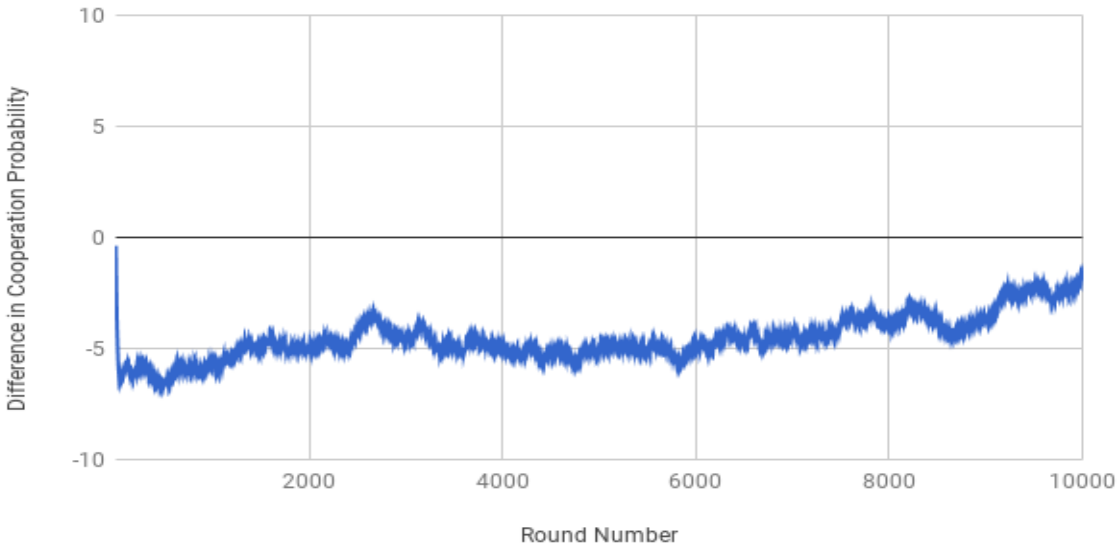

Figure 5.7: Graph of Difference of Expected and Actual Probability of Cooperation with Probability Error Adjustment For a 50 Vehicle Node Network with No Attackers (30 Runs)

they only defect while in the mix zone. From the prior analysis, the strategy made of only defecting is optimal in maximizing location privacy loss of the system when the threat of detection is not there. Therefore, with this simulation, it is not as important to analyze the effect on the location privacy as it is to determine what the maximum deviation of the difference in probability that occurs from always defecting is.

Upon analyzing a simulation run as shown in figure 5.8, the effect of adding attackers to the system can be clearly seen. There is a clear negative run in the difference in cooperation probability ending at approximately -82 in difference. Comparing to the last simulation without attackers, one attacker has made an increased deviation from a zero average to 0.8. Referring to figure 5.9 and figure 5.10, the number of attackers has increased to 3 and 5 respectively. In other words, the deviation of the cooperation probability difference grows even greater.

One of the key aspects showing the change in difference in cooperation probability is to show that attackers cause a significant and visible difference in a VANET. As

a result, if attackers are not cautious in their strategy, the noise that they cause can make defenders become wary of an anomaly.

Table 5.1: Cooperation Probability Difference for 0, 1, 3 and 5 Attackers in a 50 Vehicle Node Network

Number of Attackers	Difference in Cooperation Probability
0	-2.148
1	-82.257
3	-155.106
5	-274.08

When the values for the quantity of attackers are compared as shown in Table 5.1, the deviation increase is easily recognizable. Based off these numbers, the model can start to incorporate suspicions levels into the defender nodes. Suspicion levels can be defined as detecting anomaly or possible attackers in the system based on the deviation from the zero average in cooperation probability difference. Three different levels are arbitrary defined as shown below in Table 5.2 . They are defined by the severity of suspicion levels.

Table 5.2: Defined Defender Suspicion Levels for Final Simulation

Severity	Deviation Of Average Cooperation Probability Difference Per Round
Low	0.02
Medium	0.01
High	0.005

Taking the values in the Table 5.2 to analyze with those in Table 5.1, the results are as follows:

- A highly suspicious group of defensive nodes would be able to detect a single attacker who always defect
- A medium suspicious group of defensive nodes would be able to detect when three attackers who always defect
- A low suspicion group of defensive nodes would be able to detect when five attackers who always defect

Based on the results above, in order to balance out, it is important for attackers to mix up their strategy rather than always defect.

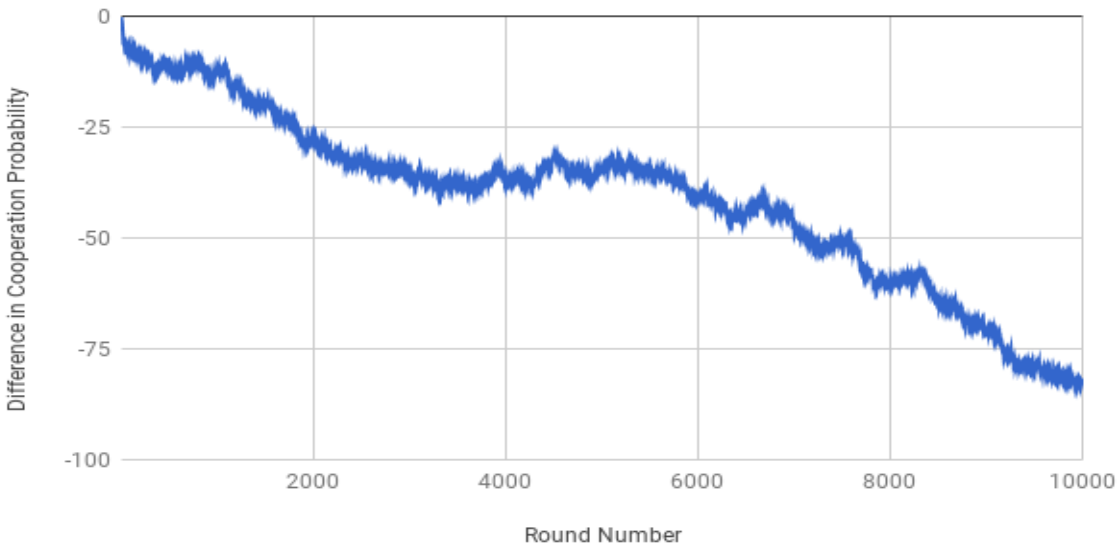

Figure 5.8: Graph of Difference of Expected and Actual Probability of Cooperation with Probability Error Adjustment in a 50 Vehicle Node Network with 1 Attacker

5.7 Final Simulation: Attackers that Remain Undetected while Maximizing Location Privacy Loss

The final simulation takes into account all other simulations that have been built up to this point. The addition to this final simulation is that there is an analysis on

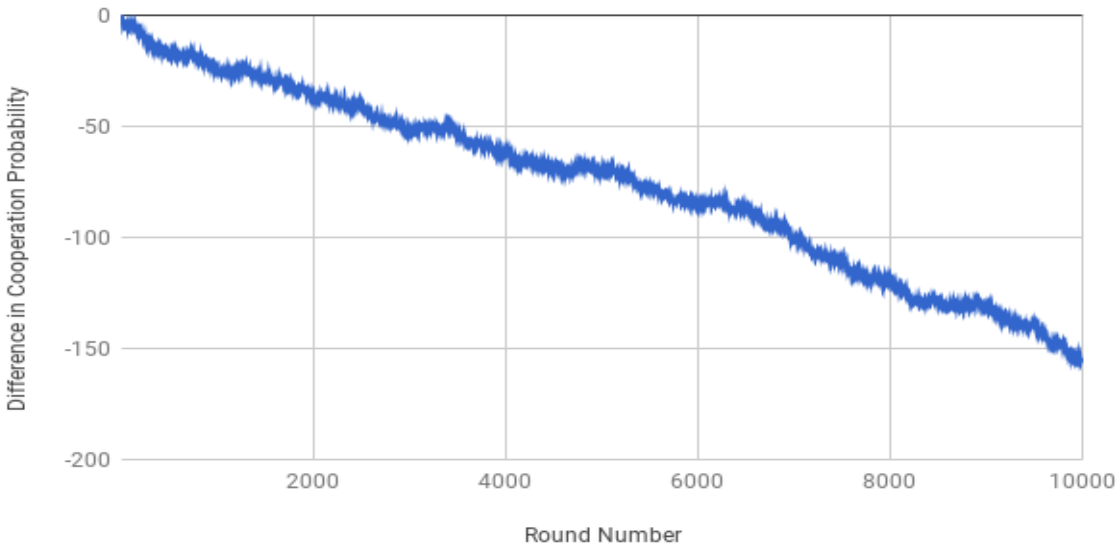

Figure 5.9: Graph of Difference of Expected and Actual Probability of Cooperation with Probability Error Adjustment in a 50 Vehicle Node Network with 3 Attackers

the attackers' success for not being detected based on Table 5.2. Additional analysis is done as well on the total location privacy loss of the system. Both of them are done for different percentage values for the rate of defect, where that percentage represents the chance for the attacker to defect when it is inside a mix zone. Finally, the model calculates an optimal strategy for the defect rate of the attacker and evaluates the affect of using that defect rate.

The optimization for this simulation is for the attacker to minimize the location privacy of the system while maximizing remaining undetected in the network. From the final game model in the design portion, the attacker achieves this by maximizing the payoff that the attacker can achieve. The difference in the optimization from the model is that the attacker also optimizes its strategy between cooperate and defect in order to achieve maximizing payoff.

This equation is expanded as follows from equation 4.7:

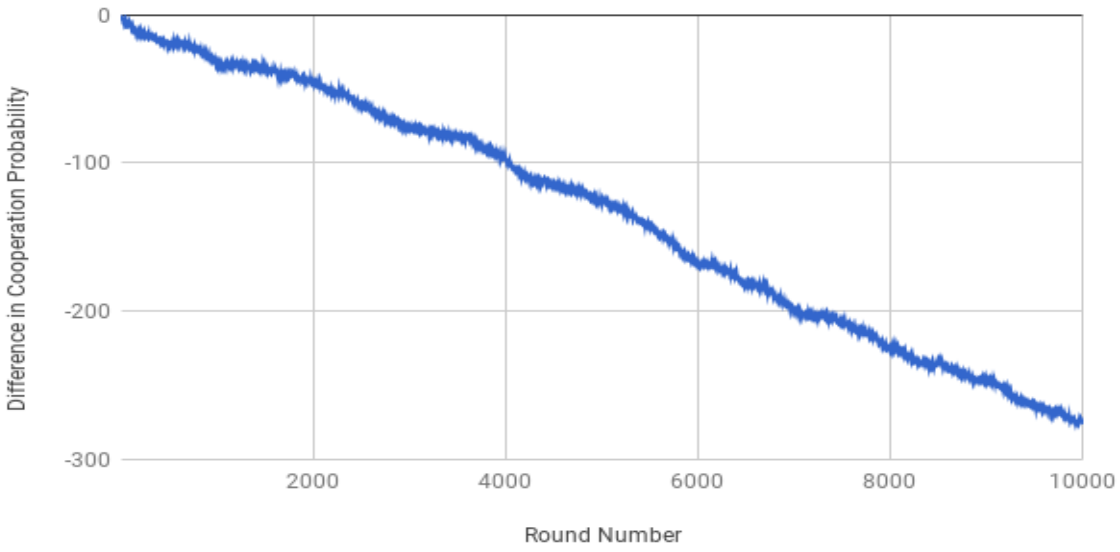

Figure 5.10: Graph of Difference of Expected and Actual Probability of Cooperation with Probability Error Adjustment in a 50 Vehicle Node Network with 5 Attackers

$$TLPL = [(n-a) * (LP_n - LP_{n-a}) - D] * (1-q) \tag{5.1}$$

Referring back to equation 4.7, in addition, q represents the probability of choosing the cooperate strategy and $1-q$ thus represents the probability of choosing the defect strategy. There is no representation of the cooperate strategy here as the payoff for choosing that is zero. From this equation 5.1, an attacker should maximize its defect rate without D, the possible detection penalty maximizing. For this particular model, D can maximize to $(n-a)*(LP_n - LP_{n-a})$ resulting to a $TLPL = 0$ when maximized.

The equation for D is as follows:

$$D = [(n-a) * (LP_n - LP_{n-a})] * [\log_{10}(10 * (1-q))] \tag{5.2}$$

Examining these equations 5.1 and 5.2, these equations are combined to obtain an equation to maximize these requirements. The equation is as follows:

$$(1 - \log_{10}(10 * (1-q)) * (1-q) \tag{5.3}$$

Table 5.3: Constant Values Used in the Final Simulation

Variable	Value
The total number of nodes in the VANET	50
The minimum number of nodes that enter a mix zone	5
The number of rounds of mix zones	10,000
The cost to switch pseudonyms	0.3
The chance to enter the mix zone	0.1
The location privacy loss per round	0.1

When finding the maximum point of this equation, it equates to a cooperation rate for the attacker of approximately 63 percent. That means that the supposed optimal defect rate for an attacker for maximizing location privacy loss and minimizing detecting is approximately 37 percent.

The results provided in the following graphs have the following constant values used in the simulations provided in Table 5.3.

Furthermore, each simulation of 10,000 mix zone rounds is ran 100 times in order to gather sufficient data. The following graphs in Figure 5.11, Figure 5.12, and Figure 5.13 demonstrate the percentage of undetected simulations for a given defect rate, suspicion level, and number of attackers.

The following graph in Figure 5.14 demonstrates the average total location privacy loss for a given defect rate and number of attackers.

Last, the simulation is ran under the supposed optimal defect rate of 37 percent. Running the simulation 100 times yields the following average effect of attackers in the system as a whole over 10,000 rounds in Table 5.4.

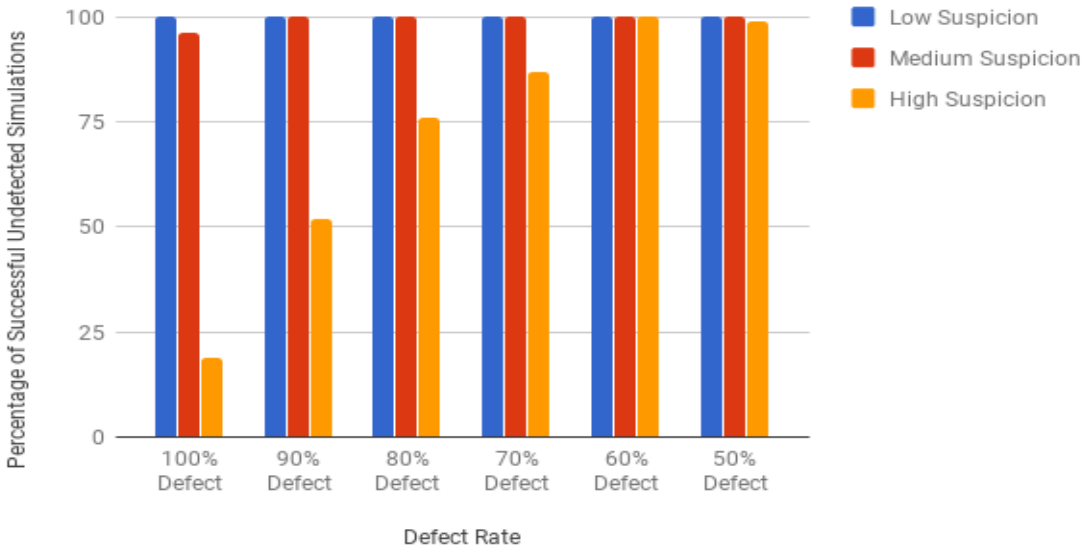

Figure 5.11: Graph of Successful Undetected Simulations Rate for Attackers for Different Defect Rates in a 50 Vehicle Node Network with 1 Attacker

Table 5.4: Effect of 1, 3 and 5 Attackers on Location Privacy with Optimal Defect Rate of 0.37 in a 50 Vehicle Node Network

Number of Attackers	Effect on Location Privacy
1	0.6%
3	1.7%
5	2.6%

5.8 Final Formal Game Model

With the results of the attacker's optimal strategy, the formal game model from the design section (Figure 4.4) can be expanded to include the mixed strategy values for the attacker. While the values of q are solved, the payoffs are still dependent on the n vehicles in the mix zone at the time as well as the number of vehicles that decide to cooperate. The basic model of the optimal strategy through the implementations

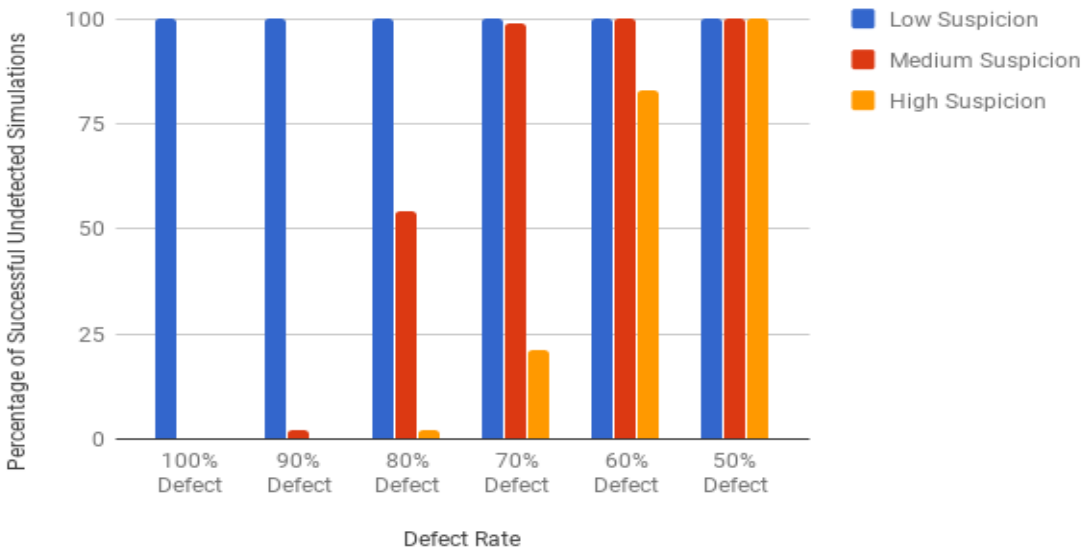

Figure 5.12: Graph of Successful Undetected Simulations Rate for Attackers for Different Defect Rates in a 50 Vehicle Node Network with 3 Attackers

previously are described in Figure 5.15. Since all of the possible scenarios of the combination of vehicles in a network that enter a network and the strategy each vehicle decides varies across each round, the model keeps the payoffs as variables from the design.

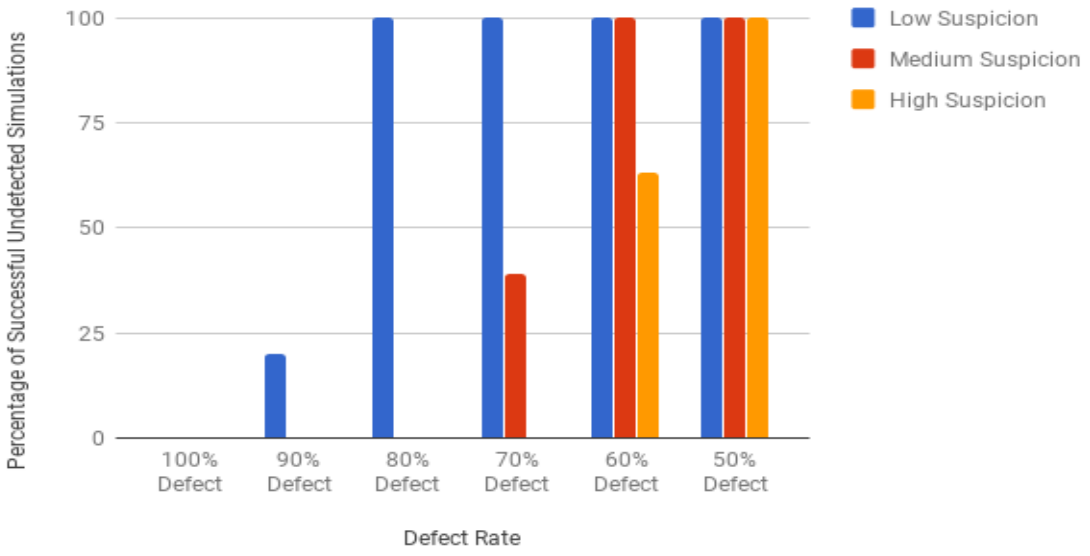

Figure 5.13: Graph of Successful Undetected Simulations Rate for Attackers for Different Defect Rates in a 50 Vehicle Node Network with 5 Attackers

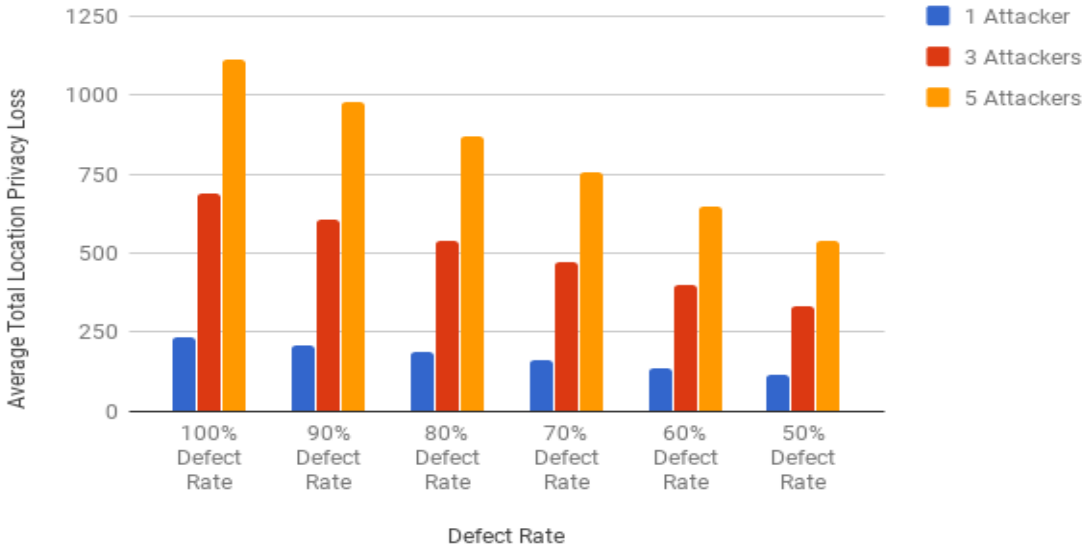

Figure 5.14: Total Location Privacy Loss Caused by Attackers in the Entire Network in a 50 Vehicle Node Network Averaged Over 100 Simulations

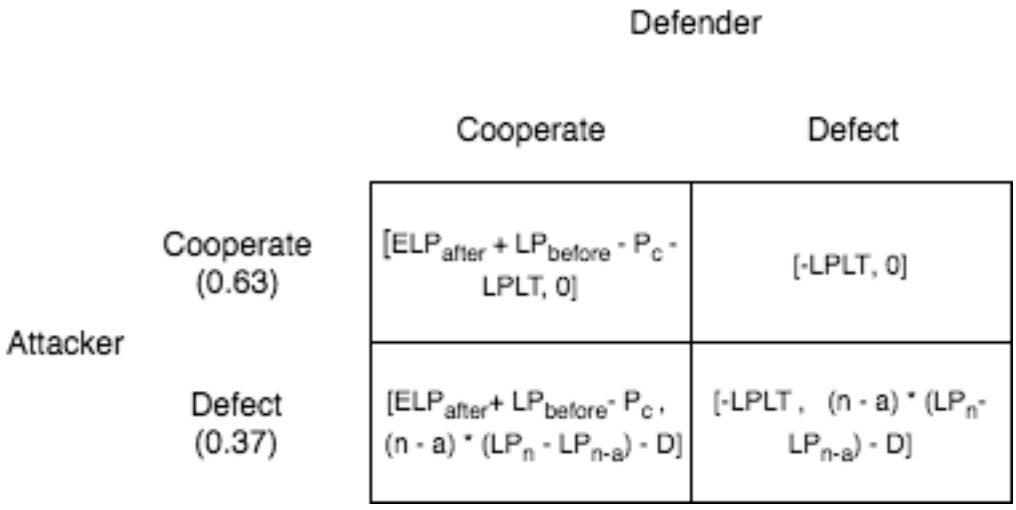

Figure 5.15: Formal Game Model for Defenders and Attackers. Formal Game Model for Defenders and Attackers. 2-player Game with Cooperate and Defect Strategies for Both Attackers and Defenders and Their Corresponding Payoffs. Includes Optimal Mixed Strategy for Attackers of Cooperate = 0.63 and Defect = 0.37.

Chapter 6

DISCUSSION

This chapter discusses the most significant results that were obtained from the implementation and simulation of the models.

6.1 Analysis of Detection Algorithm

The use the expected probability of cooperation deviation provides an excellent way to propose a way for defending vehicles to monitor the VANET system. From the results, it is clear that higher percentages of defect rates severely impacts the success rate of attackers remaining undetected. As the number of attackers increase inside a system, it becomes much harder for them to stay undetected at a higher defect rate. For a group of defenders who have a low level of suspicion (a 2 percent deviation from zero in expected cooperation probability), a single attacker or three attackers could defect every single time and maintain a very high success rate on remaining undetected. Although when 10 percent of the VANET is filled with attackers, the results show that the attackers are most likely to be undetected even at a 80 percent defect rate. In a network where the nodes may be more wary of possibly malicious behavior that represents high suspicion, the success of staying undetected doesn't become viable unless you have a 60 percent or lower defect rate. Overall, the possibility of the nodes having different levels of suspicion can play a huge role in the attackers' choice of strategy.

6.2 Analysis of Maximizing the Total Location Privacy Loss

As mentioned previously and supported by the results, higher levels of defect rate will provide the maximum payoff for attackers if they do not care about being detected. However, when the goal of remaining undetected becomes a necessity for the attackers, they have to consider the previous results from the detection algorithm in order to maintain the highest level of being undetected while still maximizing the amount of location privacy loss they obtain. Looking at the results, it seems that one of the best strategies is to maintain a defect rate that has close to 100 percent success in being undetected. If an attacker has access to know the suspicion level and current deviation of probability, the attacker could manipulate the defect rate, depending on the current scenario, in order to maintain high level of location privacy loss while staying undetected.

6.3 Effect of Attackers on the System as a Whole

When taking into account how these types of internal attacking vehicles effect the location privacy as a whole, the results show that the overall location privacy loss accounts for 0.6 to 3 percent of the total location privacy gained by the defenders. It is fairly significant when looking at numerous rounds of mix zones in large portions and find that such decrease of 0.6 to 3 percent impacts every normal user in the network. Another important statistic to take note of is that for every time an attacker defects in a mix zone, it generally takes away 12 percent of the location privacy that could have been gained if the attacker would have cooperated.

6.4 The Optimal Attacker Strategy

When looking at the calculated value of 37% for the defect rate for an attacker, there are many elements that the defect rate can not consider. First, the punishment of detection for an attacker can vary in necessity. If an attacker wants to guarantee never being detected by the defenders, the design for payoffs becomes extremely hard to solve for this game. This is because of the random nature of VANETs with mix zones and the uncertainty of what will occur in one round. Most of the design for the game comes from predictive formulas for the behaviors of vehicle nodes in the system. The predictive formulas can only analyze information that are available to all nodes where as the actual decisions that nodes make may use additional information at their disposal like current personal location privacy. Because of the difference in predictive and actual decision models, there would be no way to guarantee being undetected based on variance.

Looking at the graphs from the final model, seeing that the undetected success rate only starts "looking" guaranteed once the defect rate hits 50% or lower, this coincides with how the 37% defect rate is calculated to be optimal. The formula for calculating the optimal defect rate based on detection penalty is based on keeping the variance of suspicion as close to zero as possible. If an attacker would be able to determine the amount of variance in difference in probability of expected cooperation with actual probability that defenders have before they are suspicious, then the optimal defect rate would be able to increase. Along with this, if the attackers define the cost of being detected is not a priority, this also would raise the optimal defect rate.

Chapter 7

FUTURE WORK

This chapter discusses some possible researches om the future that could expand the concepts discussed in this thesis in order to enhance security in VANETs.

7.1 Continued Research On Threat Models On Mix Zones

Previous research on mix zones have identified adversarial threats based on eavesdropping stations in attempt to track locations of vehicles even after the vehicles switch their pseudonyms through mix zones. This thesis has identified the potential attack surface of an undetectable attacker vehicle which has blended in with other selfish vehicles. Continuing research to find more attack surfaces on mix zones or VANETs will help to further understand the impact of each attack on the system and to create defenses for those high severity attacks.

7.2 Continued Research On Detection Methods

As described in this thesis, one possible way to detect a malicious activity is to monitor the predictive model for any abnormal deviation from what is expected. Further research could analyze what types of information from broadcasts and from vehicles we can monitor to determine whether or not someone in the system is trying to act in a malicious manner. The more detection methods we can use, the more enhanced the VANET security will be in terms of monitoring the system real-time for any threats and being able to mitigate any critical threats from occurring later on.

7.3 Implementing the Model on a Real Life System

To further test the validity of the model and the implementation concepts on real vehicles in a VANET can help solidify evidence of possible real world attack surfaces. Along this, expanded scaling of the model to encompass a bigger network can help provide insight into the concept of scaling defenses and detection on a large network.

Chapter 8

CONCLUSION

Ensuring that security research is being upheld for a technologically advancing field like VANETs is extremely important in order maintain the privacy and safety of society. This thesis talks about the potential attack surface of an internal adversarial vehicle within a build security environment of mix zones.

Using past research on how mix zones have been modeled based on game theory, the paper presents its own model and implementation of an attacker vehicle that intends to maximize the amount of location privacy loss.

From the results that were found through the analysis, there is strong evidence that these internal attacker vehicles can have a significant impact on decreasing the location privacy of a VANET, even with a small number of attackers. As mix zones often support very large vehicle networks, it is important that there is a way to determine whether there is malicious adversarial action occurring in a mix zone.

Overall, this thesis helps modeling and evaluating an attack surface for which future security research can establish defenses in order to maintain the safety and privacy of the drivers and passengers in VANETs.

BIBLIOGRAPHY

[1] Cal Poly Github. `http://www.github.com/CalPoly`.

[2] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, Jan 2003.

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

[4] L. Bindschaedler*, M. Jadliwala*, I. Bilogrevic, I. Aad, P. Ginzboorg, V. Niemi, and J.-P. Hubaux. Track me if you can: On the effectiveness of context-based identifier changes in deployed mobile networks. *Proceedings of the 19th Annual Network amp; Distributed System Security Symposium (NDSS 2012)*, 2012.

[5] A. Boualouache, S. M. Senouci, and S. Moussaoui. A survey on pseudonym changing strategies for vehicular ad-hoc networks. *IEEE Communications Surveys Tutorials*, PP(99):1–1, 2017.

[6] W. A. Chaab, M. Ismail, M. A. Altahrawi, H. Mahdi, and N. Ramli. Efficient rate adaptation algorithm in high-dense vehicular ad hoc network. In *2017 IEEE 13th Malaysia International Conference on Communications (MICC)*, pages 23–28, Nov 2017.

[7] S. K. Chowdhury and M. Sen. Attacks and mitigation techniques on mobile ad hoc network x2014; a survey. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, pages 11–18, May 2017.

[8] B. DeBruhl, S. Weerakkody, B. Sinopoli, and P. Tague. Is your commute driving you crazy?: A study of misbehavior in vehicular platoons. In *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, WiSec '15, pages 22:1–22:11, New York, NY, USA, 2015. ACM.

[9] F. Dötzer. Privacy issues in vehicular ad hoc networks. In *Proceedings of the 5th International Conference on Privacy Enhancing Technologies*, PET'05, pages 197–209, Berlin, Heidelberg, 2006. Springer-Verlag.

[10] S. Du, X. Li, J. Du, and H. Zhu. An attack-and-defence game for security assessment in vehicular ad hoc networks. *Peer-to-Peer Networking and Applications*, 7(3):215–228, Sep 2014.

[11] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. C. Parkes. On non-cooperative location privacy: A game-theoretic analysis. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, CCS '09, pages 324–337, New York, NY, USA, 2009. ACM.

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

[13] M. Humbert, M. H. Manshaei, and J. Freudiger. Tracking games in mobile networks.

[14] L. Lamport, R. Shostak, and M. Pease. The byzantine generals problem. *ACM Trans. Program. Lang. Syst.*, 4(3):382–401, July 1982.

[15] A. Luckshetty, S. Dontal, S. Tangade, and S. S. Manvi. A survey: Comparative study of applications, attacks, security and privacy in vanets. In *2016*

International Conference on Communication and Signal Processing (ICCSP), pages 1594–1598, April 2016.

[16] F. Qu, Z. Wu, F. Y. Wang, and W. Cho. A security and privacy review of vanets. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):2985–2996, Dec 2015.

[17] M. Raya and J.-P. Hubaux. The security of vehicular ad hoc networks. In *Proceedings of the 3rd ACM Workshop on Security of Ad Hoc and Sensor Networks*, SASN '05, pages 11–21, New York, NY, USA, 2005. ACM.

[18] Y. Wang and I. W. H. Ho. On-road feature detection and fountain-coded data dissemination in vehicular ad-hoc networks. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–6, Oct 2017.

[19] Y. Wang, F. R. Yu, M. Huang, A. Boukerche, and T. Chen. Securing vehicular ad hoc networks with mean field game theory. In *Proceedings of the Third ACM International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications*, DIVANet '13, pages 55–60, New York, NY, USA, 2013. ACM.