2018-12-17

Framework for Anonymous Secure Data Transfer in Vehicular Ad-Hoc Networks

Jabari Stegall
Worcester Polytechnic Institute, jcstegall@wpi.edu

Worcester Polytechnic Institute
Electrical & Computer Engineering

Framework for Anonymous Secure Data Transfer in Vehicular Ad-Hoc Networks

Jabari Stegall

A Thesis Submitted to the Faculty of the WORCESTER POLYTECHNIC INSTITUTE in partial fulfillment of the requirements for the Degree of Master of Science in Electrical and Computer Engineering
by

_____

December 2018

Approved:
Professor Alexander Wyglinski, Advisor

_____

Professor Yarkin Doroz

_____

Professor Krishna Kumar Venkatasubramanian

_____

Abstract

With the increasing number of Vehicular Autonomous Network (VANET) architectures and applications, user privacy must be addressed and protected. Internet of Things (IoT) and their applications take care of everyday mundane task in order to increase user convenience and productivity. However, studies have shown that IoT architectures can be a weak spot in network security, including data being sent plain text. In this thesis, a VANET architecture is proposed that is capable of securing anonymous data collection from a distributed set of autonomous vehicles. The proposed architecture features a hybrid combination of centralized and decentralized routing concepts. Unlike other VANET implementations, our proposed architecture provides anonymity to users in the network. Lower latency can be achieved by merging data from live short range adhoc routing methods with the data collected from a pseudo live long range centralized routing methods. The proposed architecture guarantees user anonymity within the VANET framework. Most VANET models assume users do not value the privacy of their identity. We assume that each vehicle is equipped with a VANET computer capable of storing data, performing calculations, and both sending and receiving data wirelessly. Therefore vehicles can communicate directly with each other and exchange data within short distances as well as communicate with long range wireless infrastructure. Simulation results show the implementation is equipped to handle diverse traffic scenarios as well as deter adversaries to the network from maliciously trying to manipulate collected data.

Acknowledgements

I would like to express Gratitude to:

- My Advisor, Alexander Wyglinski for his guidance and support.
- Gates Millennium Scholars Program for primarily financially supporting my educational studies
- WPI Robert S. Parks Fellowship for its supplemental financial support
- My lab mates in the Wireless Innovation Laboratory for their support and encouragement
- Finally my Wife, Sister, Mom, & Dad for their love and support

"Try not to become a man of success, but rather try to become a man of value."

-Albert Einstein

"Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid."

-Albert Einstein

"Gonna eat till Im tired and then sleep till Im hungry."

-Deadpool

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Objectives

The Insurance Institute for Highway Safety Highway Loss Data Institute reported statistics for fatal vehicle crashes in 2016. The report states that there were $34,439$ fatal car accidents resulting in $37,461$ deaths [33]. The benefits of autonomous (*no human intervention required to operate*) and semi-autonomous (*some human intervention required*) vehicles are mostly hypothetical since they have not been deployed on a large scale. Self driving vehicles have the potential to significantly reduce fatalities by intervening when human error is detected, such as collision detection, with features such as automated breaking in semi-autonomous vehicles. In the near future, it is expected that autonomous vehicles can potentially remove human error all together by communicating location, speed, direction, and traffic jams to each other with V2V networking infrastructure. Autonomous vehicles can also receive road information from driver-operated semi autonomous vehicles and vice versa.

Data shared between vehicles could improve safety, increase travel time efficiency, and contribute to the facilitation of large scale autonomous vehicle deployment [1]. These vehicles will transmit and receive a vast amount of data, which will have to be processed and stored in a database. Local

databases in vehicles will be fundamentally important to all future VANET architectures since it will allow vehicles to make decisions in real time across a variety of real road scenarios [1]. While vehicles collect data, they could also send it to a global database that covers a specified area, such as a city or certain group of subscribers. The vehicles within a coverage area or subscriber group could potentially receive traffic data otherwise unavailable to their location in order to enable autonomous and/or semi-autonomous vehicle operations with drivers in order to avoid possible accidents or other adverse traffic incidents. While surveying the literature with respect to VANETs and Internet of Things most proposed IoT applications assumed users will not mind divulging their identity during data collection to further the progression of IoT applications such as VANETs. This assumption is not based upon reliable polling data, therefore an alternative infrastructure we proposed that assumes users want privacy. This infrastructure should:

- Provide a way for users to submit data to a VANET while protecting their identity,
- Collect and verify useful user data while preventing submission abuse,
- Propagate useful verified data to a VANET in a timely and practical manner.

1.2 Current State of the Art

A popular V2V implementation is combining Dedicated Short-Range Communication (DSRC) [22, 21] with Global Positioning Satellite (GPS) [22, 21] technology, such as the model described by the National Highway Traffic Safety Administration (NHTSA) [22, 21]. Their implementation optimizes local ad hoc communication amongst nodes and offer reliability, security, positioning accuracy, and ease of installation. Centralized approaches via a line-of-sight satellite communication is great for connecting vehicles in different locals. However, satellite communication is LOS dependent and has latency issues. It could benefit by having the vehicles possessing local ad hoc communications until satellite signal returns. Reference [2] proposed an implementation

that possesses a decentralized approach that relies on high node density for ad hoc communication. However, this implementation lacks the ability to communicate with vehicles that fall out of the communications network. Nevertheless, the vehicle that is out of range should be able to communicate to a centralized database that has data collected from the vehicles within the ad-hoc network.

A hurdle in trying to design a VANET network is trying to figure out how to structure a network with constantly moving nodes. Applying modern routing methodologies whose behavior is implemented in a centralized manner and combining it with decentralized routing methodology can create robust information centric network [4, 6, 7]. Vehicles that are in a low vehicle density area can receive VANET data from long distances from the centralized arm while the real time V2V data can provide the most up to data information in high vehicle density situations.

1.3 Research Contributions

In this thesis, a framework is proposed that utilizes decentralized and centralized vehicle networking in parallel for optimal and timely data collection This hybrid implementation leverages the VANET computer capable of storing data, performing calculations, and both sending and receiving data wirelessly, has GPS information, and can perform processes and make decisions from local processing. The VANET computer receives data from neighboring vehicle nodes via short range communication as well as from a centralized data providers such as a municipality that collects and propagates information from all nodes in the specified coverage area such as the city limits. The short range communication enables the peer-to-peer ad hoc communication that provides real-time vehicle information. The long range communications rely on data from a centralized data collector/provider using information collected from vehicles sensor data. Long range communications ensure that vehicles located outside of the peer-to-peer network web range can receive accurate road information in a timely manner. Anonymity of subscribers in this architecture is ensured

because the vehicles only send non identifying information status/sensor data, location data and time stamps in respect to those locations, and various other event data collected. Since that data verifier receives data using a different dynamic header than the data collector, the verifier and collector are hindered from working together to identify users by cross referencing user data. The subscribers are also protected from the manufacturers of the VANET computers working from with the data collector and verifier to identify users because the serial numbers of the machines are not used in data transmission. The primary data attributes are location of instance and time of instance. Referencing information by time instances allow V2V and centralized database data to co-mingle in the calculations to enable more robust vehicular decision making. This redundancy provides great latency and coherence by providing autonomous and human operated vehicles with temporally and spatially relevant data.

1.4 Thesis Organization

This thesis discusses the prerequisite concepts necessary to understand the discussed later in the Chapter 2. Chapter 4 will consist of the proposed system model of the hybrid topology and it will be discussed and described in detail.Chapter 3 will further discuss the control variable "Relevance" $R(t)$ and it will be initially be defined in a general sense then modified to display the network operators management, control, and intervention capability. Chapter 3 continues to describe a proposed implementation of the centralized topology that provides secure anonymous data collection to protect user information using various methods will be explained and analyzed in detail. Possible methods of a attack that could adversely effect the behavior of the network will be considered. The decentralized ad hoc smart information centric routing option will also be proposed and described. Lastly, simulations and results from attack scenarios in the centralized topology as well as results surveying Routing quality in the decentralized topology in a discrete event space will be presented and discussed.

Chapter 2

Vehicle Network Security Overview

This chapter provides the background information needed to understand the chapters that follow. It delivers a basic overview of a Vehicular Autonomous Network detailing the infrastructures, nodes, the motivation of VANETs as well as briefly discuss the standards, routing methods and security implemented. Secondly, this chapter investigates blind signatures as a method of anonymity. Finally, it briefly discusses the centralized and decentralized network topologies as well as wireless ad-hoc networks used in the implementation of the hybrid system.

2.1 Vehicular Autonomous Networks VANETS

Vehicular communication is defined as the communication between the vehicles. The main objective of deploying a VANET is to reduce the level of accidents [16]. The general premise of a VANET is to provide Intelligent Transportation System (ITS) [18] services to end users. This service is useful for providing information such as safety and traffic information via fast data exchanges with other vehicle nodes in the network or various VANET infrastructure nodes. It uses different standards such as DSRC and WAVE for fast data communication. Many routing protocols have been designed for implementation of routing in VANET [16]. Nowadays, researchers are focusing

on designing secure VANET systems to prevent them from different malicious drivers who disrupt the network performance [1, 2, 5]. VANETs can be affected by active attacks such as Denial of Service[1] that attempts to congest network traffic by consuming all the network bandwidth. VANETs are also susceptible to passive attacks such as "man in the middle" where a third party attempts to intercept communication between nodes. With these method of attacks and others in mind, VANET routing protocols must be developed to protect from these attacks. VANET provide many services to the end users such as multimedia sharing, content delivery, security, and e-health facilities [17]. VANET researchers are working on issues such as routing, broadcasting, security, traffic management [18, 19], and information fusion [19] and so on.

VANET architecture mainly consist of three node types moving Vehicular Nodes, stationary Rode Side Units (RSU), and various VANET infrastructure that supports the operation of the networks [1, 16]. The vehicles sending and receiving data in the network primarily serve as nodes for data transmission in the VANET. However, the RSU can act as a router and if strategically placed it can provide better network coverage for vehicles in the network. Communication is performed Vehicle to Infrastructure (V2I), Vehicle to Vehicle (V2V), or Vehicle to Infrastructure to Vehicle (V2I2V). In order for the vehicles to communicate in the network, they must be equipped with a on-board communication systems capable of sending and receiving data as on-board system should also include GPS capability as the bare minimum for VANET implementation. Additional add-ons suggested by [18] is an Electric license plate (ELP), which can serve as a unique identifier of sorts. The ELP implies the users do not mind if their identities are available on the network. Later in this paper, we propose a method to protect user identity in the VANET. A Certification Authority (CA) exists in the architecture for providing services, applications, and managing the live network. This paper later discusses how to work with a CA without compromising user identity. You can see the architecture of a VANET in Figure 2.1. Intelligent Transportation Systems [18] imply that a vehicle in the network itself acts as a sender, receiver and router for broadcasting information. This ITS must address and regulate communication for V2V and V2I communications. Vehicles communicate with each other in two types of ways. The first is receiving broadcast that are either

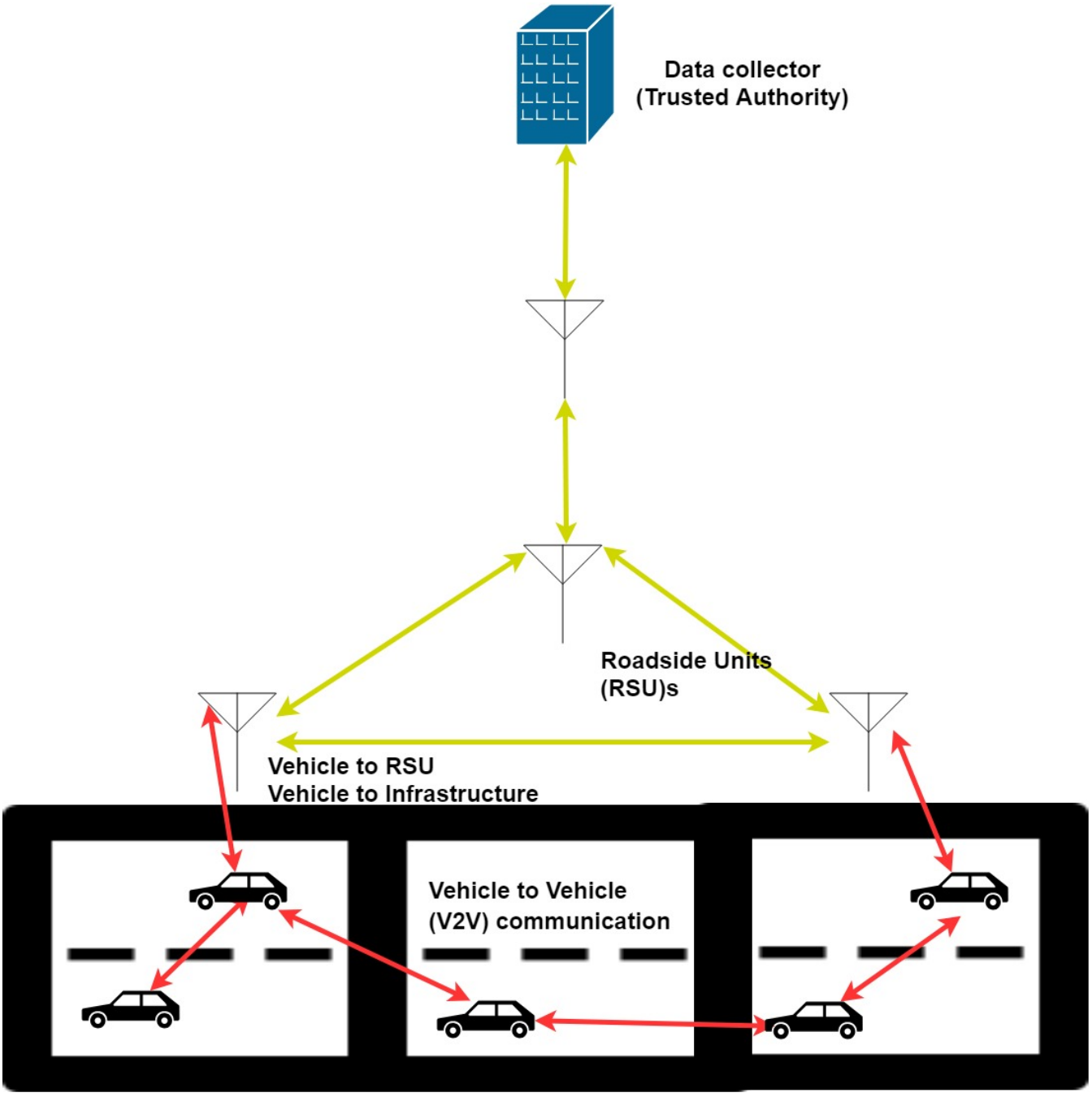


Figure 2.1: VANET architecture depicting vehicles communicating V2V, V2V, and V2I2V.

sent in regular intervals, which can cause message collisions, or received broadcast emergency messages that have a higher routing priority due to some particular importance. The type of information that would be sent in regular intervals should be information that will not drastically affect the network if there are minor lapses in packet delivery such as vehicle velocity, direction, location, and emergency messages. That information can be used to help vehicles in the VANET identify traffic events such as car accidents or help vehicles using GPS find better routes. Messages with high priority that can be pushed to the network are necessary because they can be used in emergency situations to alert vehicles in the VANET of possible seriously adverse situations, such as inclement weather or messages alerting the approach of a emergency vehicles with lights on. The second method of communication for vehicles are on demand requested. This intelligent broadcast be implemented using a ranking systems that determines if the information would be interesting or useful to the end recipient [5, 6]. Sending targeted broadcast is a way of reducing the collision domain of data transmission. The RSU are also useful in reducing collisions and increasing bandwidth in the VANET by serving as a reliable stationary node in the network.

2.1.1 VANET Standards

Standards are used for the development of the product as well as to assist users to verify and compare the products. Many standards are employed according to the protocols used to define network behavior such as security, routing, and services. Currently the popular standards used in VANETs are dedicated short range communication (DSRC) [20] and wireless access in vehicular environment (WAVE) [21, 22]. DSRC was designed to support a variety of applications based on vehicular communication [20].It provides short to medium range communication. The primary motivation for deploying DSRC is to enable collision prevention applications and these applications depend on frequent data exchanges among vehicles, and between vehicles and roadside infrastructure [20]. The US Federal Communication Commission (FCC) allocated 75 MHz of spectrum at 5.9 MHz for DSRC, which consist of seven channels. DSRC-based collision avoidance aims to increase safety by

using information received from DSRC neighbors such as location, speed acceleration, and state information. This information is transmitted with a high duty cycle and can be used for vehicles to provide warnings, alerts, and make smart decisions. DSRC operates on the physical layer of the OSI model while the a standard called WAVE operates on the MAC layer. These standards are complementary to each other for successful VANET operation.

2.1.2 Routing in VANETs

Routing data in a VANET can prove difficult due to the high mobility of nodes in a network. Most routing methodologies are built around the assumption of ad-hoc environments. The main issues in VANET which require routing are network management, traffic management, broadcasting, mobility, topological change, quality of service, and fast data transfer and so on [16]. In order to address those many impedances, a sophisticated routing techniques must be developed. Routing protocols are divided as such: topology based, position based, cluster based, geo cast based and broadcast based [23]. For the purposes of this thesis, topology, position, and cluster will be discussed.

Topology-based routing is the most identifiable form of routing as it is applied to most telecommunication networks. Topology based routing [23, 24] is further divided into proactive and reactive subsets. In a proactive routing methodology, a routing table is already established so path discovery is not necessary. However, maintaining low traffic or unused routes can lead to high network load and excessive bandwidth utilization, which will lead to the degradation of network performance. Some examples of proactive routing include, Destination Sequenced Distance-Vector Routing [23], OLSR: Optimized Link State Routing Protocol [23], and Cluster Head Gateway Switch Routing [23] to name a few.

Due to proactive routing protocols having load and bandwidth consumption problems, reactive routing protocols are implemented by making route discovery on demand. Therefore, network

load is significantly reduced due to optimized route management wasting less resources. Examples of protocols that use this methodology to achieve low packet overhead are, Dynamic Source Routing [23], Ad Hoc on Demand Distance Vector [23], and Junction-based Adaptive Reactive Routing [23]. Hybrid routing protocols are being designed to take advantage of the strengths of both proactive and reactive methodologies. Proactive routing can be used for small scale micro routing while the large scale macro routing can be executed use reactive methods.

2.2 Cryptography Techniques for Communication

One of the primary motivations for the proposed framework is providing anonymity to users in the VANET while still allowing the collecting of useful data to allow the network to thrive. It will be assumed that someone always will and can listen to messages transmitted between nodes. However, if transmissions are properly encrypted the "man in the middle" [10, 11, 25] will receive nothing more than indiscernible information without the cryptographic key. However, a method to verify or vet data must still be in place to maintain the integrity of the the network. In network architectures sensitive information could be the identity of communicants, the contents of messages sent between communicants, or a combination of identity and message contents.

Public key cryptography can allow an electronic mail system or network packet forwarding system to hide the content of the communication despite an unsecured underlying telecommunication system [11]. With this method you can send keys in unsecured channels to whoever is interested in your key. For instance, correspondent A can send a message correspondent B's public key. Since the message was sent using the public key of correspondent B, correspondent B's private key is the only way to decrypt the message. Public Key Cryptography was the solution to providing both parties in communication with a secret communication key [25, 26]. You can view a example of the system model in Figure 2.2

Secure communication is important when sending message through mediums such as email. In

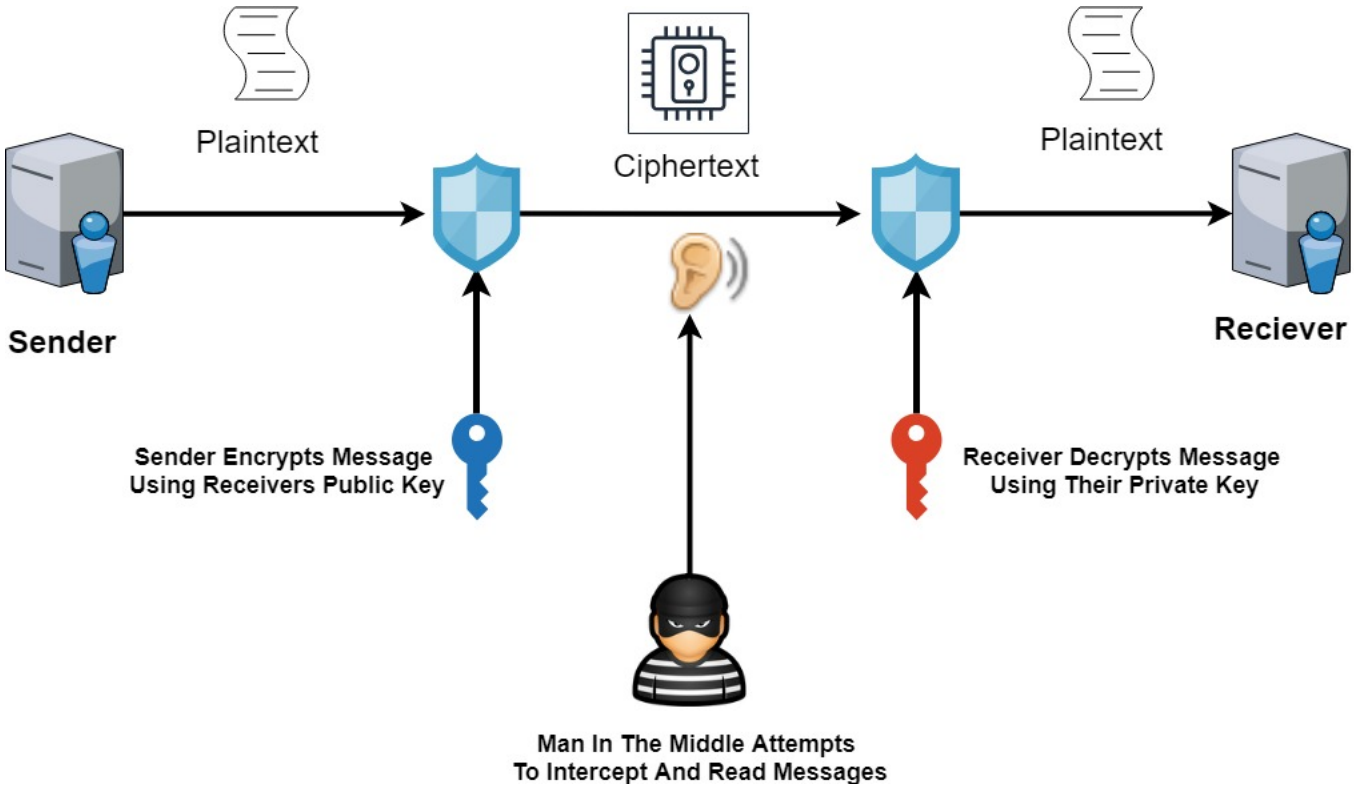


Figure 2.2: Public Key Algorithm System Model depicting how data is encrypted with a public key and later decrypted with a private key. Data encrypted with a private key cannot be deciphered by third parties without the private key.

email applications, the public key is the email address while the private key is the password that a user uses to log into their private mailbox. The algorithm for public key cryptography is shown in Algorithm 1.

Algorithm 1 Public Key Encryption

1: **procedure** INPUT(Msg)
2: **Encrypt** $En(Msg, Pubkey)$ ▷ Encrypt Msg with receiver Public Key
3: *Send Encrypted Msg through communication channel*
4: **Decrypt** $(En(Msg, Pubkey), PrivKey)$ ▷ Decrypt Msg with receiver Private Key
5: **Output** (Msg)

In Algorithm 1, the user creates a message with the intended receiver's public key. The sender transmits a message through a semi secure channel that is visible by some third parties if they are listening. The message is encapsulated by encryption protecting the privacy of the message content. Once the receiver intercepts the message they decrypt the message sent to them using their private key, which is the only way the contents of the message can be deciphered. This methodology is useful for securing messages. However, it would not be feasible in a VANET as encryption and decryption is calculation intensive and would increase the overall latency in a network. Therefore, a method with less computational overhead is necessary.

In VANETs, large amounts of sensor and vehicle status data is collected. Vehicle location, electronic identification, direction, and velocity are often values generated and transmitted to other vehicles or infrastructure in the network. That data can be processed locally in the vehicle's computer/on-board communication. Data can be processed by a central data collector and retransmitted to vehicles. The data received from the central data collector will have more a delay in accuracy than data collected and process real-time at the vehicle. You can use data fusion to merge data from real-time data collection and data collected from a database. While modern VANET implementations are created with the assumption that users do not desire anonymity, it does not consider many users who will not adopt VANET technology due to lack of privacy. user identity is not necessary for data calculations or decision making.User identity is more for verification to sustain network integrity. Since vehicles in a VANET architecture share their location

frequently, vehicles can can make informed decisions based their a neighboring nodes broadcasted information. Considering these factors, there needs to be a method in place to collect verified user data while maintaining user anonymity. Chaum [9] provides an example of blind signatures that could be implemented for verified financial transactions, which aims to inhibit third parties from determining the payee in a transaction, the time the transaction was made, and the amount of the transaction. Furthermore, individual should have the ability to provide proof of payment as well as determine the identity of an payee in unprecedented incidences. The last feature is the ability to stop the use of payment media such as credit cards or checks that have been reported stolen or suspected to be compromised. The algorithm for blind signatures is displayed in Algorithm 2

Algorithm 2 Blind Signatures

```
1: procedure INPUT(x)                ▷ x is chosen at random such that r(x), and forms c(x)
2:     Sender (Send c(x) to C.A. )                                  ▷ C.A.- Certificate Authority
3:     C.A. (Sign c(x))                                  ▷ c(x) becomes s'(c(x)) after signed by C.A.
4:     Sender (Send s'(c(x)) to Sender )
5:     Sender (Strip s'(c(x)) with c' )      ▷ After stripping s'(c(x)) the sender is left with s'(x)
6:       if (s(s'(x)) ≠ x)
7:         Discard s'(x)
8:       else
9:         RETURN
10:    Sender (Check r(s(s'(x))) = r(x))                ▷ Perform validity check before sending
11:      if (true)
12:        Send to desired receipient
13:      else
14:        Discard s'(x)
15:    Receiver (Check r(s(s'(x))) = r(x))              ▷ Perform validity check before processing
16:      if (true)
17:        Process data
18:      else
19:        Discard s'(x)
```

This section gave a brief overview of the methodologies for public key cryptography and blind verified signatures. The motivation for public key cryptography is to provide correspondents with the means to send communication keys in public unsecured channel. Blind signatures is a cryptographic methodology that provides an environment to conduct, verifiable, anonymous, auditable financial transactions. In Chapter 3, a system i proposed that employs functions from

the public key and blind signature methodology to facilitate verifiable anonymous data collection in VANETS.

2.3 Centralized and Decentralized Network Topologies

A Network topology can be defined as the arrangement of interconnected nodes in a communication network [27]. The physical topology describes the geographical location of communication nodes in a network. The logical topology is characterization of the flow of data in the network. The logical topology is usually identical to the physical topology. However, this is not the case with a vehicular autonomous networks as the vehicle nodes are constantly on the move creating a unpredictable physical topology.

Centralized topologies are exactly what their name suggest: a computer networking model where resources and network management is facilitated in one main location. A key function a centralized network is a "routing table" which are predefined paths in a network for data to travel through. Due to the nature of centralized networks their implementation can be expensive due to unused paths that are maintained in the network. However, the abundant expensive paths provide high reliability enough though some nodes in transmission may or may not fail as well [28]. Centralized network topologies provides network mangers and administrators the ability to control and monitor the whole network by allowing better access to the devices in the network, the network resources can be managed with less effort, and also it provides control over the security and behavior of the network. In chapter 3 ,the network manager/administrator operations and how security problems that may arise in a network such as denial of service or identity spoofing will be addressed.

Contrary to centralized topologies, decentralized networks can be characterized by using peer to peer connections to route data. Data does not have to travel to a central hub before it reaches another recipient. With this methodology, your topology does not have to have fixed points that suites a VANET well for communication. Large scale Internet applications such as VANETs can

benefit from an ability to predict round-trip times to other hosts without having to contact them first [31, 32]. In a VANET, network management is conducted by the nodes in a network such as the vehicle's in it's on-board computer with network capability as well as road side units. Since messages will be broadcasted then rebroadcasted over multiple hops to the nearest neighbor, packet loss will be less likely to occur high vehicle density situations due message overlap [29]. While the redundant messages increases the likelihood the message will be received by neighbors who will find the information useful, unnecessary storage resources will be consumed. Chapter 4 discusses how data management will be optimized in the proposed framework. The network manager has much less overhead because the responsibility of the network manager is reduced to setting use parameters such as the number of messages sent per second or address network anomalies and attacks. Network transmission information is generated located at the nodes as well as local network management.

2.4 Chapter Summary

V2V communications is a method that could revolutionize vehicle safety and travel efficiency. Just like any other network, it is susceptible to attack by malicious parties and the framework must defend against various scenarios. Vehicles use each other as routers to send information to other vehicle nodes in the network. Infrastructure nodes can also be used to communicate with vehicles and forward messages as well increase network coverage and bandwidth. Vehicles send general information that could be useful to other neighbors in the network so messages are sent out as broadcast to any available nodes in range. Broadcasts can incur collision and packet loss so method must be put in place to reduce those inevitabilities such as targeted broadcast on static reliable infrastructure forwarding vehicle data. The two main network standards receiving attention in the area of V2v communication is DSRC and WAVE. Network management can be difficult in constantly changing VANETs, so choosing the correct routing topology is key in having a successful robust vehicular network. proactive routing topologies since centralized topologies have

high overhead and network performance will eventually degrade. Reactive topologies eliminate the shortcoming of proactive routing but low manageability since routing is decentralized. Hybrid protocols are necessary to take advantage of the strengths of each methodology while overcoming the weaknesses. Cryptographic methods have been proposed to protect VANETs, although high overhead of such security could possible degrade network performance as well. Vehicular networks are complicated and and difficult to implement as many location specific and environmental factors must be considered before deployment.

Chapter 3

Centralized Anonymous Data Collection in VANETS

3.1 Motivation and Objectives

VANETs are the future of vehicle safety and traffic optimization. Many users will be eager to adopt the technology of the future especially if it means hands free driving and reduced vehicle accidents. In order for human operated and autonomous vehicles to function, it requires a significant amount of preinstalled data and algorithms to facilitate pilot-less driving. While the collection of this data is needed to support the VANET, many users would be wary about their identity being compromised due to the sensitive nature of the data such as locations and times at those locations. There needs to be a system to collect useful user vehicle data anonymously while also keeping the integrity of the data collected. The vehicles will need to improve their decision making capability as it experiences the road and various environmental events. In order to support the vehicles making optimal decisions, it needs the most accurate real-time data gathered from the neighbors as well as the host vehicle itself. The data gathered from users must be stored and processed before it is broadcasted to other vehicles in the network. The data collectors must have a quick efficient

process to filter data collected from users as while making sure its relevant for real-time vehicle decision making. Furthermore, data collectors must possess as a method to prevent users who have been reported or suspected to be sending fraudulent data from submitting data to the database for a time period. An entity that can represent a data collector as well as a network manager is the municipality of a city. The city limits can represent the physical location or subscription area of the network. Users with compatible vehicles that are network capable and/or autonomous driving capable that enter the city limits are possible subscribers. Assuming this subscription to the network service will cost money, users need the ability to opt out of sending or receiving vehicle data in the network to and from the provider if desired. Users need an incentive to submit data to an entity that will inherently make money off of its data. A good incentive for early adopters in data submission would be free access to data collected by the data collector for vehicle decision making. Data collected from a vehicle's local environment will help an autonomous driving capable car navigate more efficiently and safely in its current environment. Consequently users that support the operations of the VANET are rewarded for their efforts creating a healthy codependent relationship. The state government should oversee communication that lies between city limits such as state highways and act a data hop for data exiting a city limit. City municipalities should be connected to other city municipalities via the state government node in a large macro network. With that in place, users who may be traveling through subscription areas/micro networks such as a city will have seamless coverage for data that will be used to optimize smart vehicle decision making. Figure 3.1 depicts the different possible level of abstraction for this network framework. A way to add more value to the connected micro networks is the VANET capable cars having the ability to enter in a route for a trip and receive the most relevant information along the users expected route.

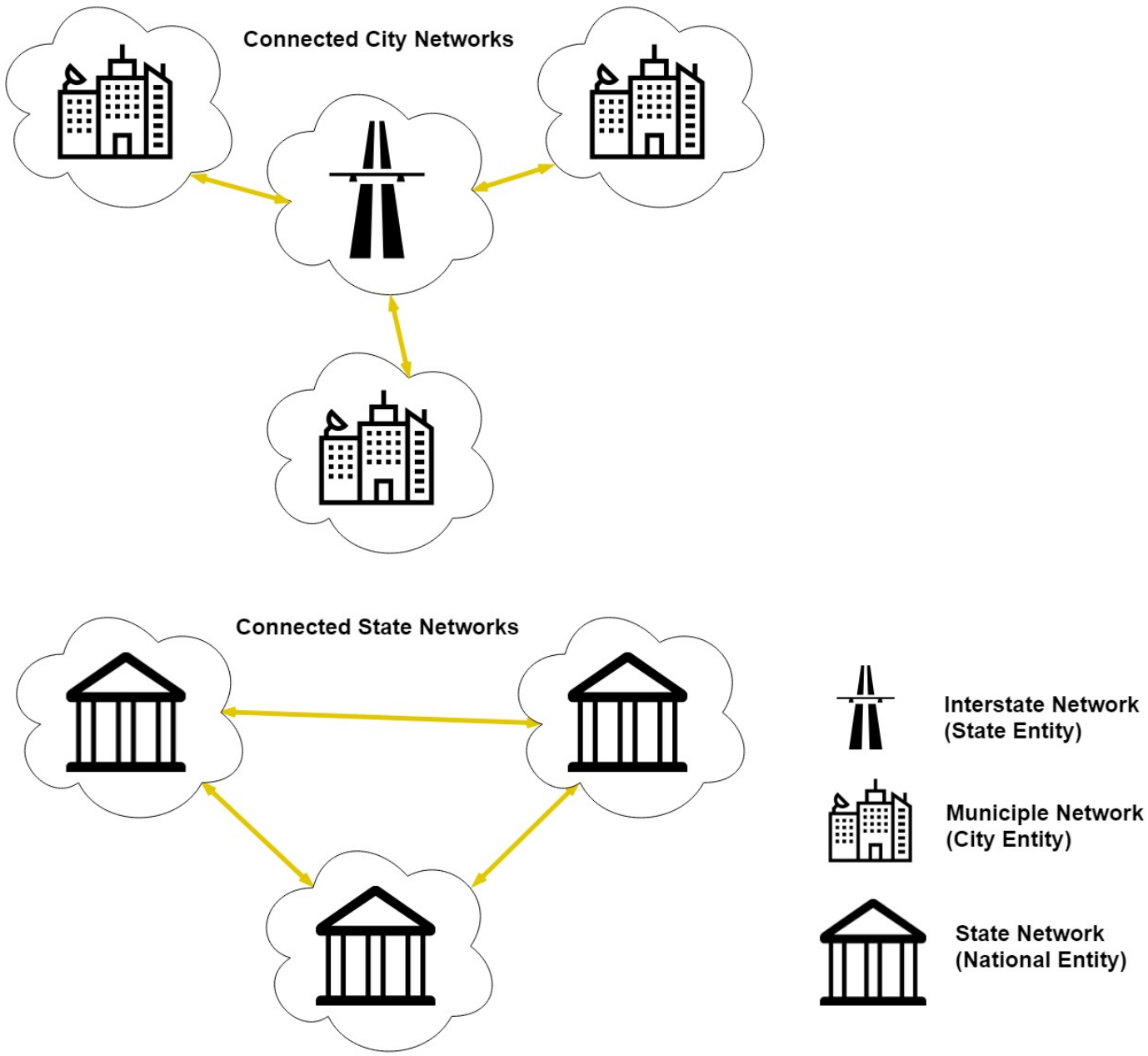


Figure 3.1: Different Abstraction levels for VANETs. City VANETS are connected by interstate entities while state networks can communicated directed to create a seamless subscriber network

3.2 Implementation & Methodology

In this thesis, we propose a hybrid centralized and decentralized routing topology for secure data transmission in a VANET to preserve user identity. Chapter 4 describes in more detail how the centralized and decentralized topologies will work together. This chapter mainly focuses on the centralized topology and explains its functions and characteristics. Relevance is a key parameter in the hybrid VANET as it can impact operations in the centralized and decentralized frameworks of the network. Relevance will initially be defined in a general sense then modified to display the network operators management capability. Relevance is designed to keep the most up to date data flowing through the network, manage network resources by purging or archiving data that has reach its expiration, as well as address network attacks by possible adversarial subscribers.

Figure 3.2 illustrates the data verification process of the proposed centralized information centric routing framework. Figure 3.3 depicts how the vehicles communicate with data collectors/verifiers in the centralized algorithm, and Figure 3.4 depicts how V2V communication is handled in the proposed decentralized V2V routing algorithm. The decentralized module of the framework will further be expanded upon in Chapter 4 so we will focus how vehicles communicate with the collectors and the verifiers. These processes work in parallel to create the proposed hybrid topology. The hybrid nodes, piloted vehicles with or without autonomous driving capability, and non piloted autonomous vehicles, are equipped with a VANET computer. The VANET computer equipped and integrated into the vehicles are assumed to have a computational ability, network connectivity via peer to peer or long range communication, as well as memory storage. Another hybrid node is a roadside unit(RSU), which sends special broadcast messages or emergency information to the piloted vehicles and non-piloted vehicles to be controlled by the network operator of a respective municipality. The RSUs can also function as hops for decentralized and centralized communication increasing network bandwidth and reliability. The four types of nodes are depicted in Figure 3.5.

Special/Emergency messages from a RSU has a higher relevance than messages or data sent by

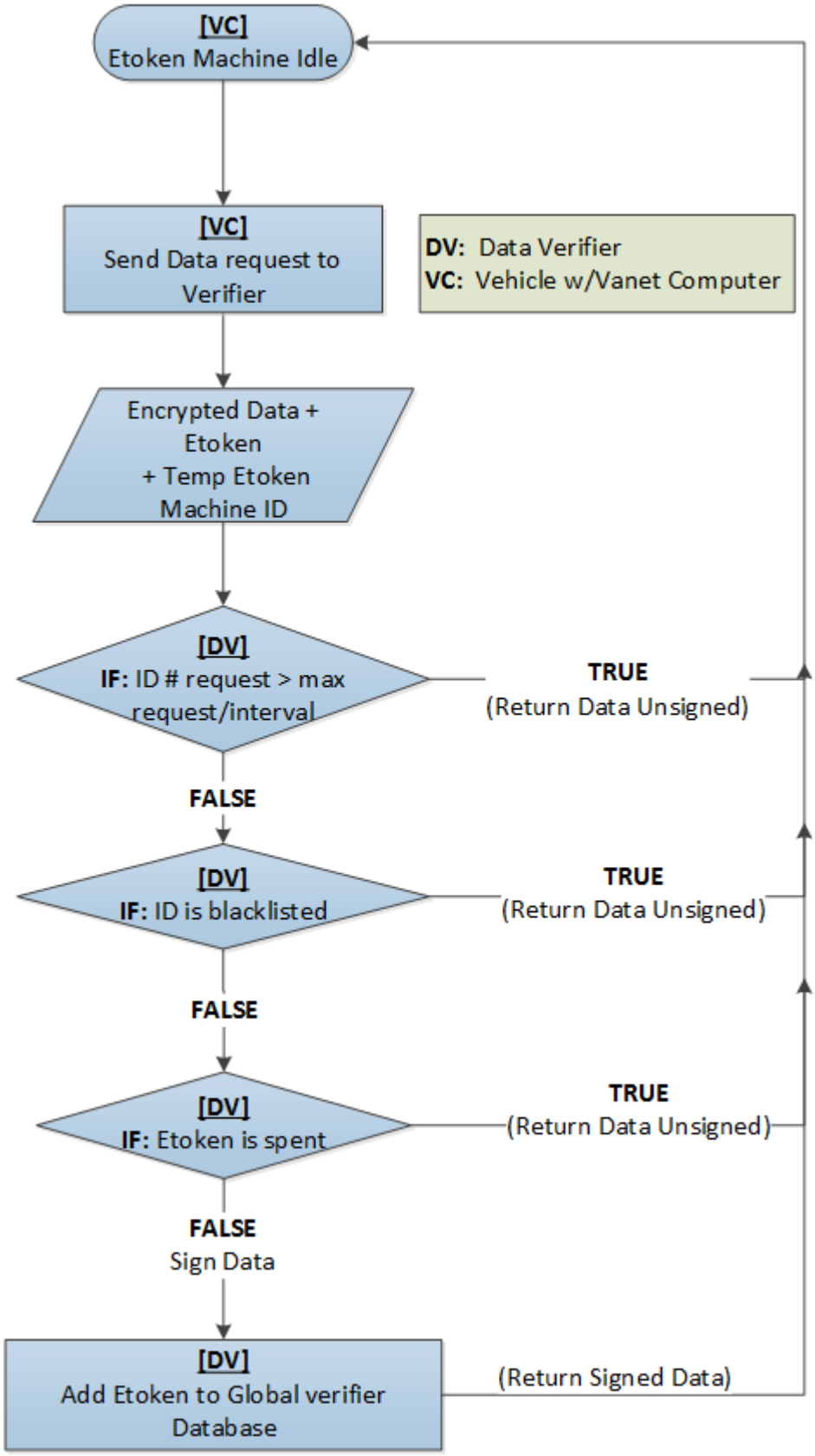


Figure 3.2: Methodology for vehicle to Data Verifier. Vehicle data to be verified must pass three test before it recieves a signature.

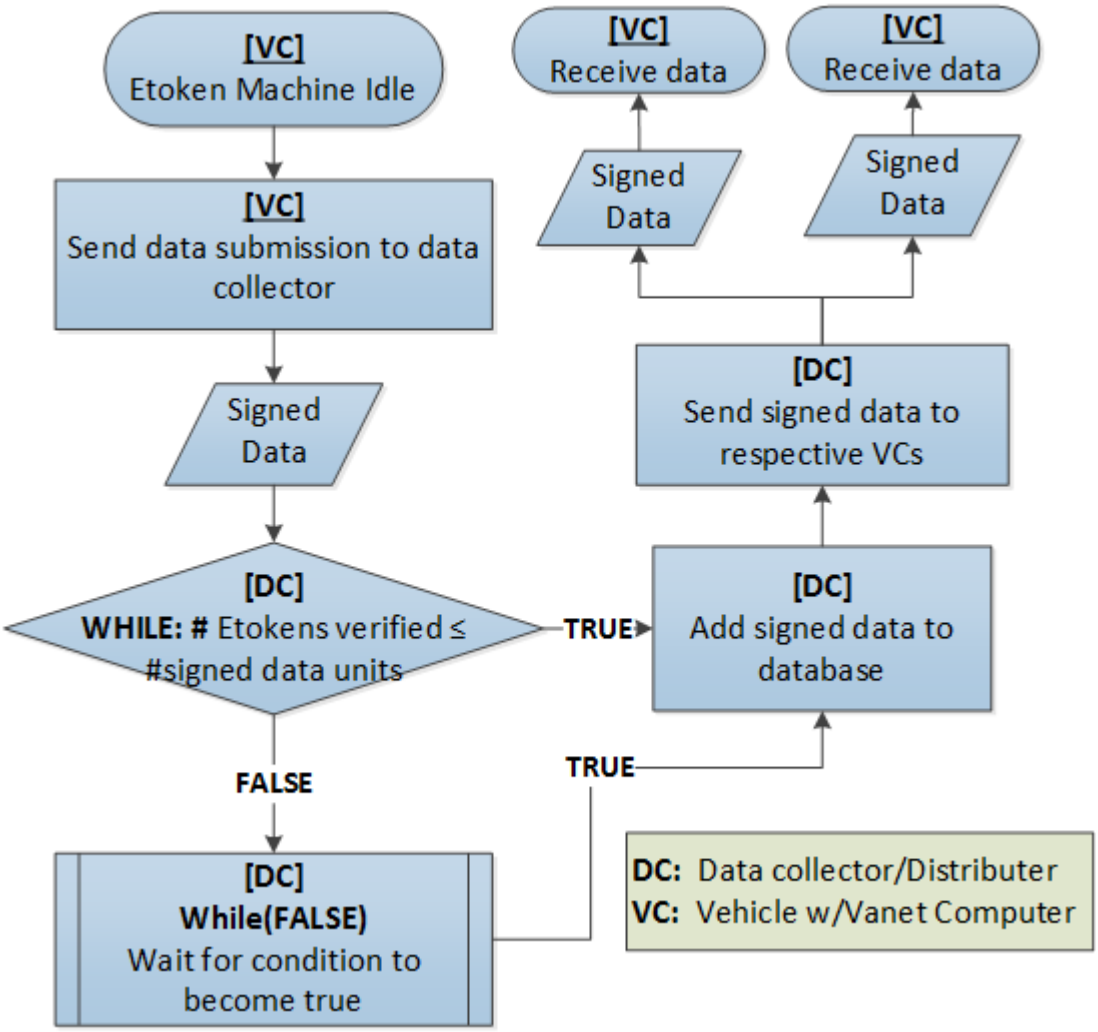


Figure 3.3: Methodology for vehicle to Data Collector. The data collector collects sign data from a vehicle in the VANET but no more data than that has been verified

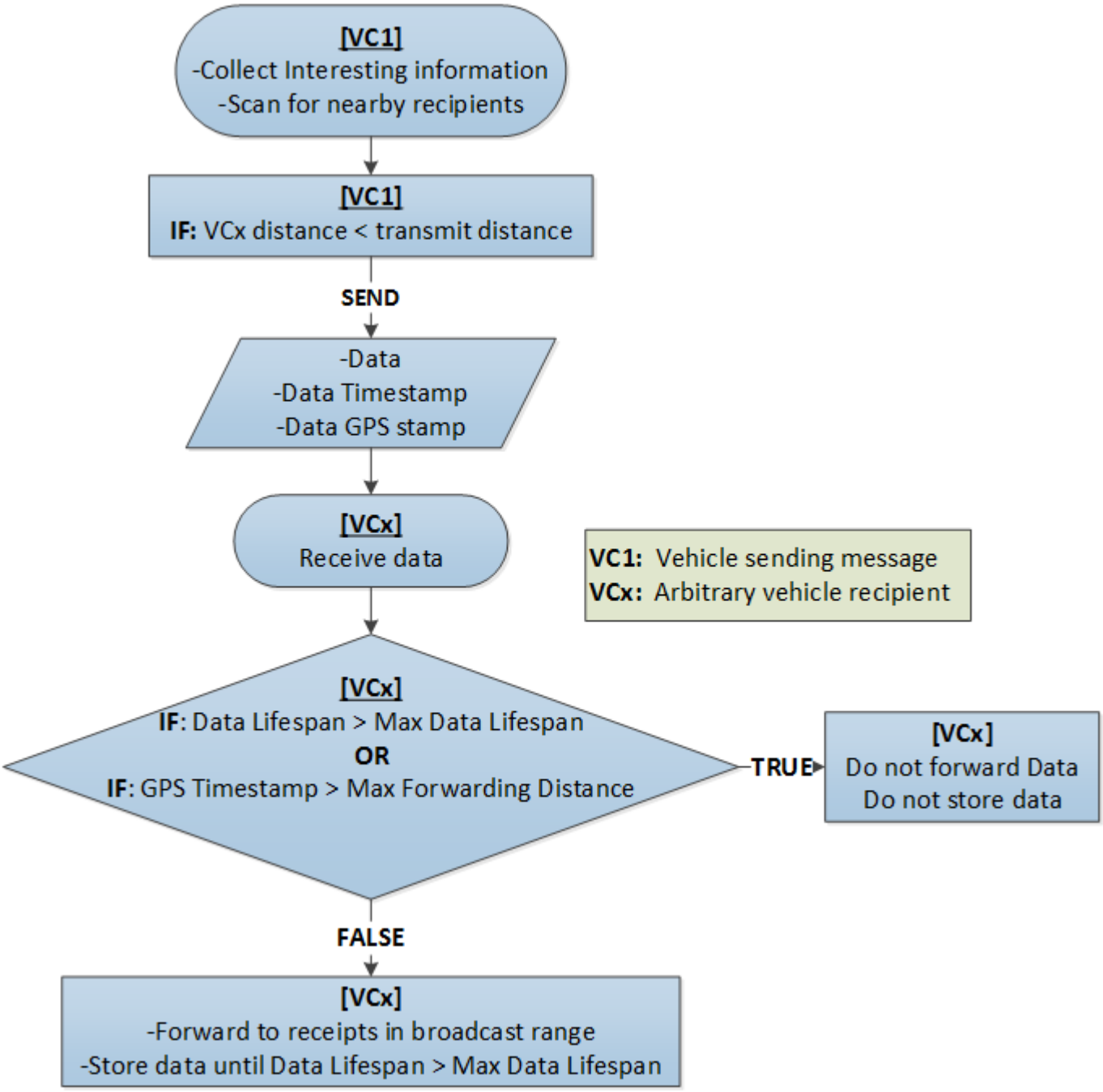


Figure 3.4: Methodology for vehicle to Vehicle. The lifespan and forwarding distance of a V2V data is controlled by spatial and temporal relevance.

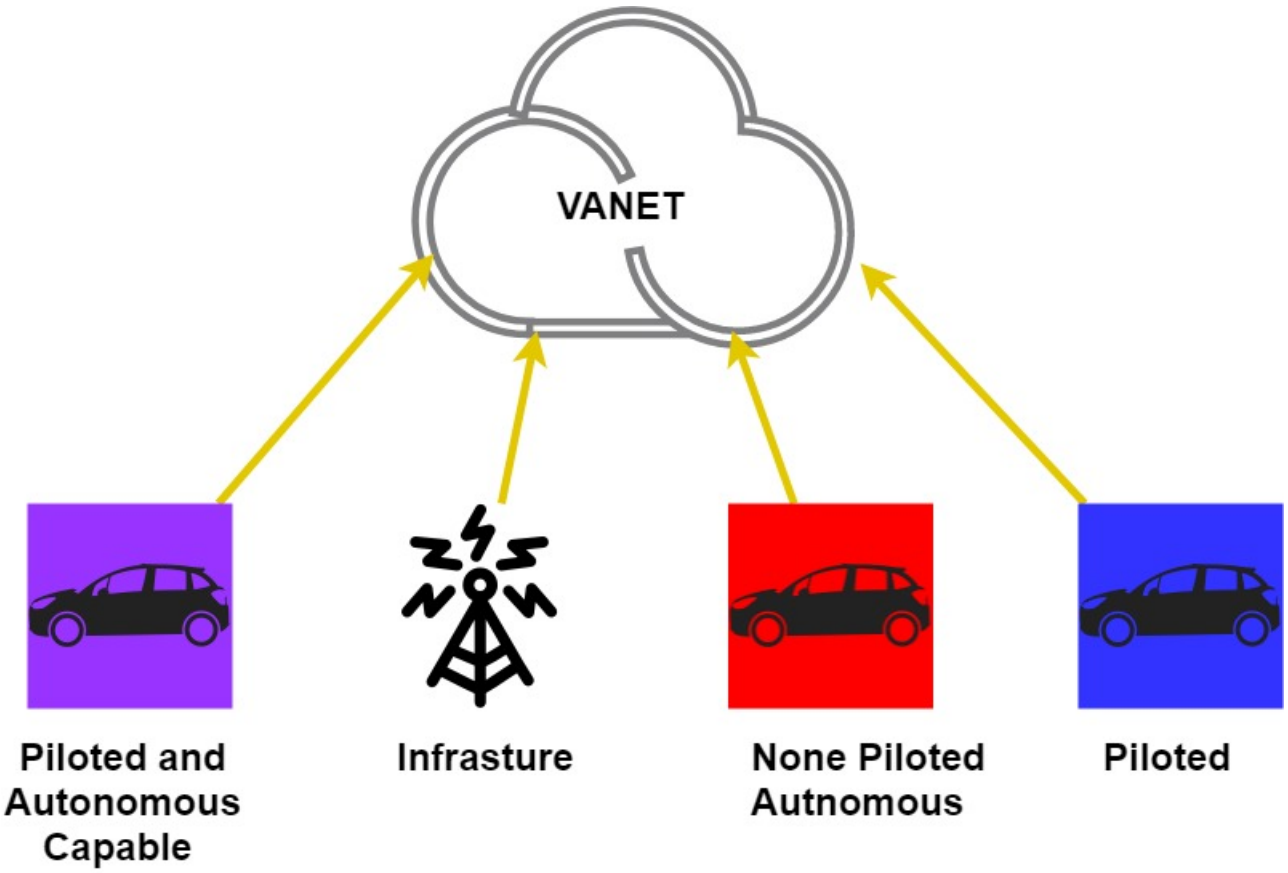


Figure 3.5: Four types of VANET nodes in a hybrid topology framework.

piloted vehicles or non-piloted vehicles. These special messages can be used by the municipality to send messages to subscribers in the subscription areas about important events or emergency situations. The relevance function is defined as follows:

$$R(t) = (t+1)e^{-t}, \quad t \geq 0 \tag{3.1}$$

At $t = 0$, relevance is equal to 100 % which implies that all information is most relevant at its instance and becomes less relevant as t, which is this case is time, approaches infinity. Equation 3.2 has been modified with the variable n. The n parameter gives the network operator control over the prioritization of the collection of data. It also protects the network from entities deemed adversaries by having the ability to set $n < 1$ in specific locations in the network, illustrated in Figure 3.6, Thus preventing other piloted vehicles or non-piloted vehicles from receiving unnecessary or false information. For regular messages from piloted vehicles and non-piloted vehicles, $n = 1$. However, for emergency broadcast or another special message $n : 2 > n > 1$. The n parameter gives the network operator control over what messages vehicles receive in emergency situations. It also protects the network from entities deemed adversaries by having the ability to set $n < 1$. Thus preventing other piloted vehicles and non-piloted vehicles from receiving unnecessary or false information.

$$R(t) = (t+1)e^{\frac{-t}{n}}, \quad t \geq 0 \tag{3.2}$$

The VANET computer has an authentication method proposed for the centralized network from [3] and inspired by the implementation in [1]. This inspired authentication is used when piloted vehicles and non-piloted vehicles send data to data collectors not peer to peer. This method is put in place for piloted vehicles and non-piloted vehicles to send verified data to data collectors while hindering data collectors from identifying the users or the owners of the vehicle, thereby allowing the data collector to sell the information to third parties or use for their own network. The subscriber can be put to ease as their personal information is not compromised. When non-piloted vehicles communicate with other non-piloted vehicles or piloted vehicles, data is intelligently routed

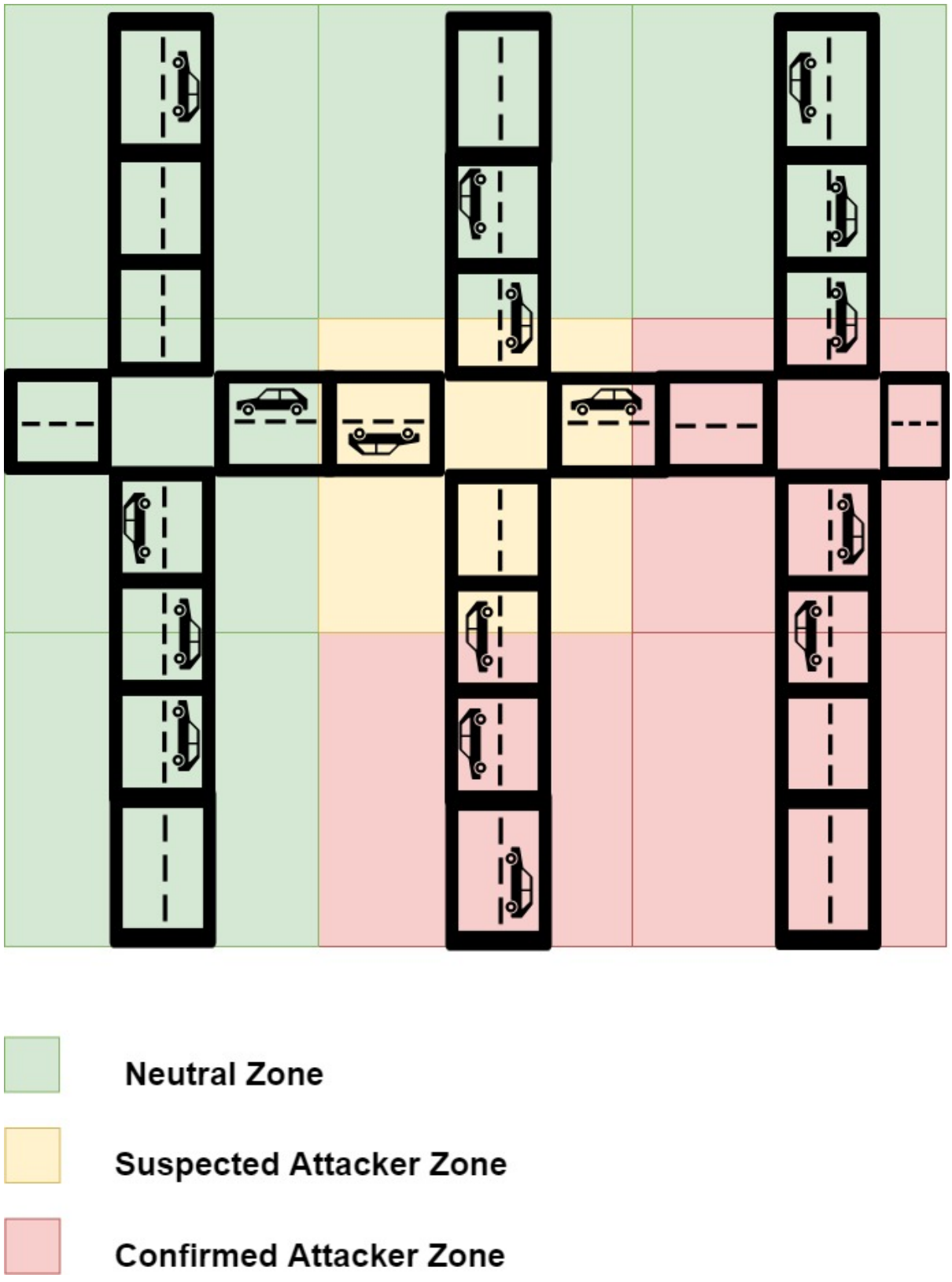


Figure 3.6: Example of network operator classifying zones in the network to modify the relevance of data.

based on the concepts discussed in [4], [5]. Information at time t will continue to be routed to other cars based on the relevance $R(t)$ and the Smart Info Routing algorithm in [6]. The VANET computer has x-times authentication control parameter set by the network operator to prevent adversaries from attempting to spam the network with authentication requests. This x is another control variable to allow the network operator to address certain anomalies in the network. Figure 3.8 is an expansion of Figure 3.6 showing how the network operator can limit the flow of data in certain zones.

In Figure 3.7, the network operator has set x to 10 in the neutral zone, 7 in the suspected attacker zone, and 2 in the confirmed attacker zone. Assuming the auditing method has sufficient accuracy, the network will limit the amount of fraudulent data entering the database. The governing body of the network should take the necessary steps to mitigate and correct the suspected attacker zones. Since there are non adversarial users in those confirmed attacker zones, the longer it takes to alleviate the cause of the high attacker zone, the more "good" data the host network in missing out on. The network security can act as a double edged sword if not properly managed and audited.

In conjunction with limiting how many messages can be submitted per interval, the network operater can change the life-cycle of messages in different zones to maintain the quality of the data collected for the database. With the current set parameters shown in Figure 3.8, the value of n in the neutral zone equals 1, in the suspected attacker zone it is equal to 0.8, and in the confirmed attacker zone is equal to 0.5. This n is from Equation (3.2). After one second of messages in the neutral zone, we will have a relevance of 0.74. With a relevance threshold set to 0.2, messages in that zone will be purged or archived at around three seconds. Messages in the suspected attacker zone will have a relevance of 0.65 after 1 second. These messages have a slightly short life-cycle compared to the messages in the neutral zone, so the possible fraudulent data does not populate the database as fast as data in green areas. The network operators perform an audit on that area to determine if it will be designated as a neutral zone again or a confirmed attacker zone. Messages in the yellow zone will be purged or archived at around 2.4 seconds with the relevance threshold of

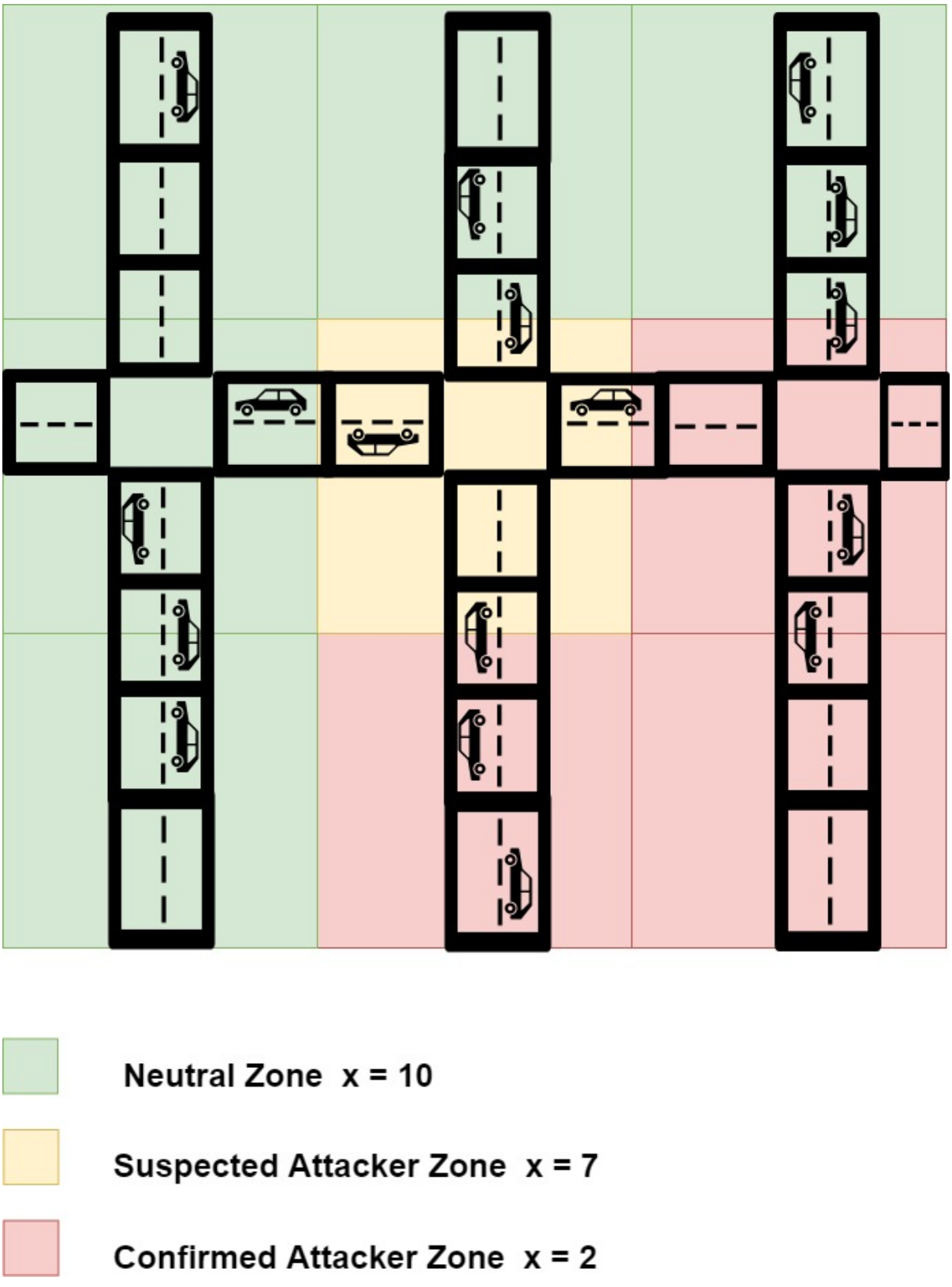


Figure 3.7: Example of network operator determining how many x messages can be submitted per submission interval.

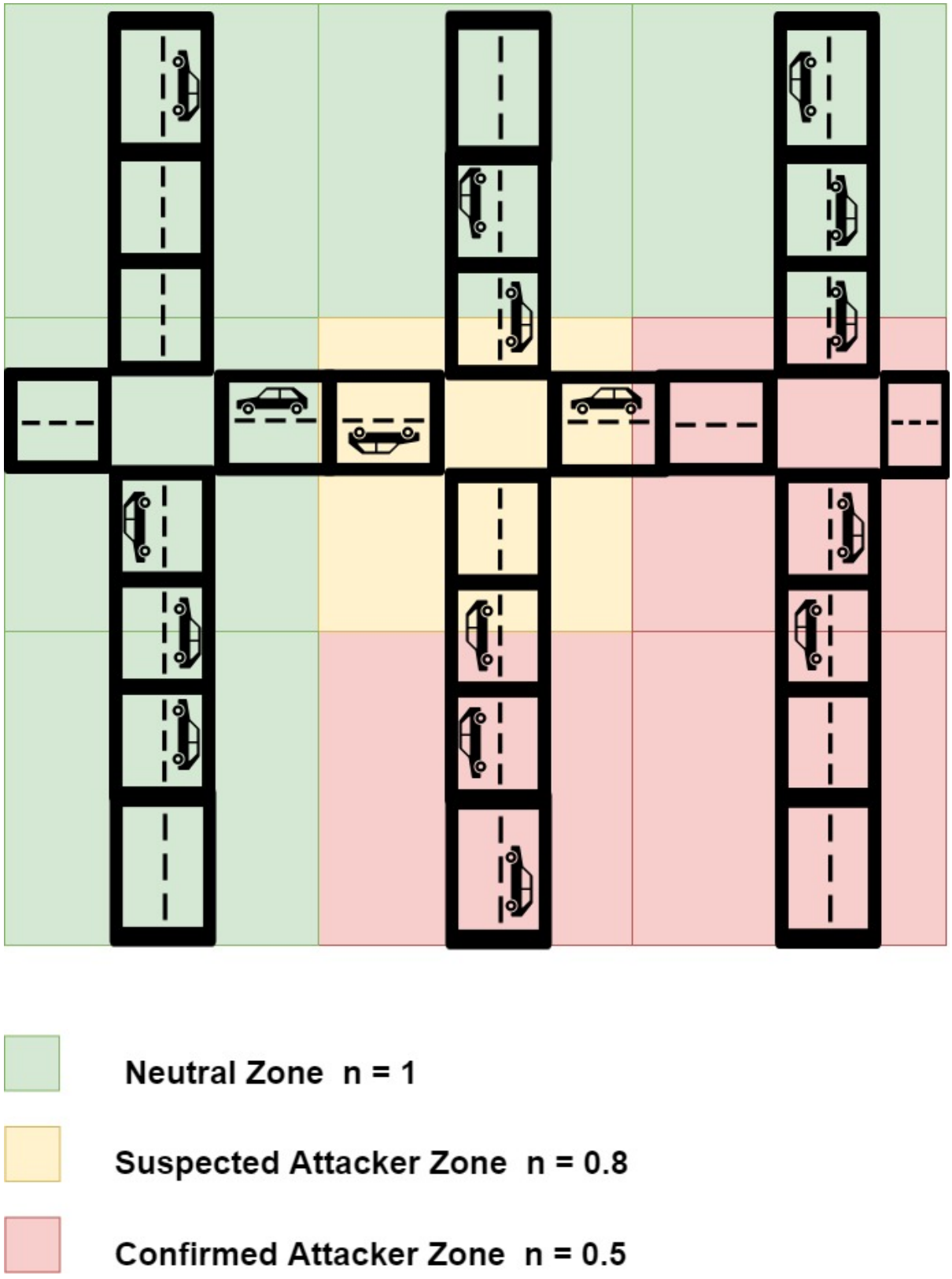
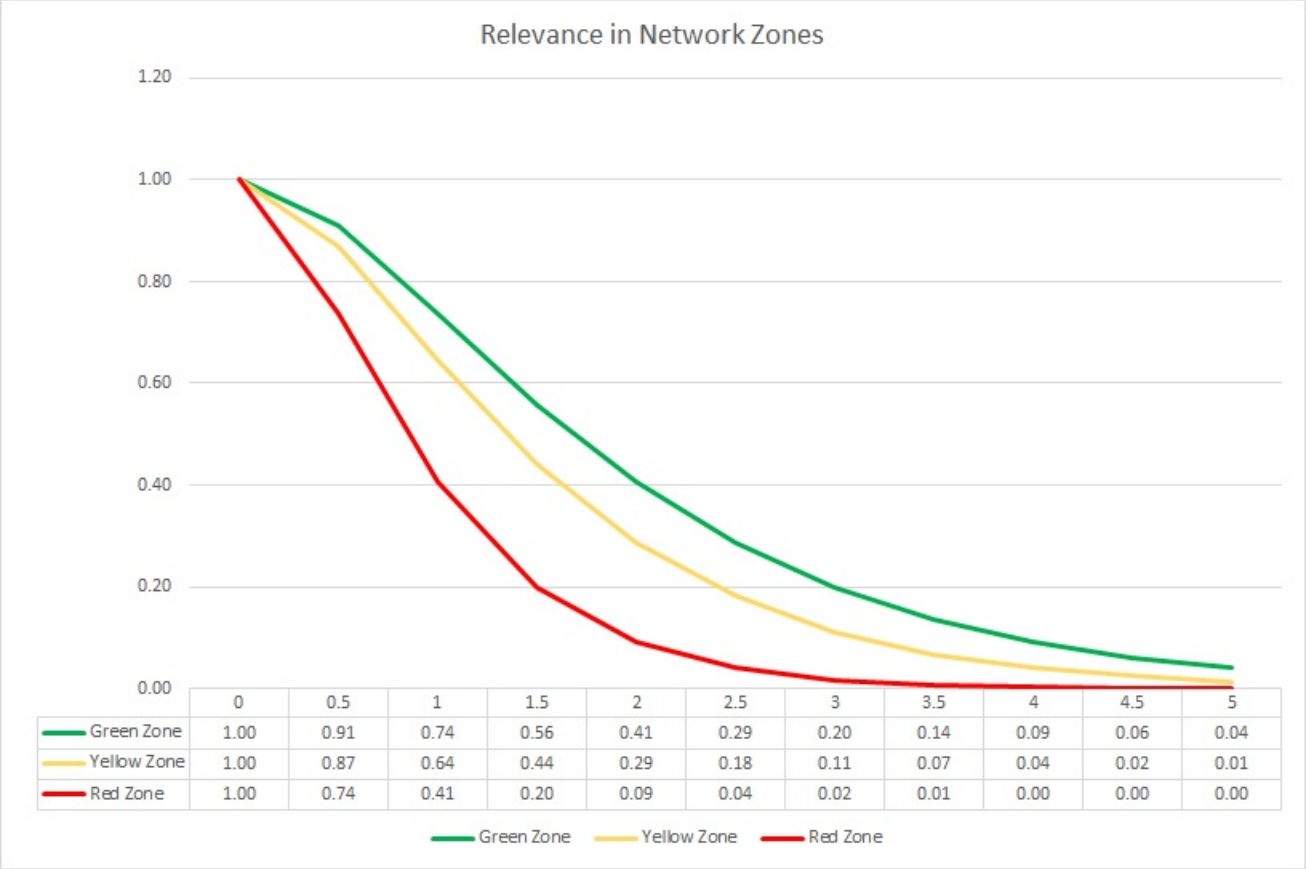


Figure 3.8: Example of network operator modifies relevance in different zones to manage the quality of data collected.

0.2. Messages in the confirmed attacker zone will have a relevance of 0.4 after one second. Zones are only labeled like this when it is determined to be a high density adversary zone. Some users in that area will be non fraudulent users, although their data will not have a meaningful contribution to the database since their broadcasted information will be purged or archived after around 1.5 seconds. The decay of the messages in each respective zone is shown in Figure 3.9



	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Green Zone	1.00	0.91	0.74	0.56	0.41	0.29	0.20	0.14	0.09	0.06	0.04
Yellow Zone	1.00	0.87	0.64	0.44	0.29	0.18	0.11	0.07	0.04	0.02	0.01
Red Zone	1.00	0.74	0.41	0.20	0.09	0.04	0.02	0.01	0.00	0.00	0.00

Figure 3.9: Relevance over time in Neutral, Suspected attacker, and Attacker zones. Relevance in red zones decay much faster in the red zone than the green and yellow zones.

This x-times authentication feature allows you to set a maximum number of requests to be sent in a chosen interval by network operator. Another feature of the VANET computer is Anonymous Proof of Authentication and Anonymous Data Transfer. These features will be further explored in the next section.

3.2.1 VANET Computer and Anonymous Data Transmission

In order to facilitate anonymous data verification, the VANET computer must have two dynamic identifiers. Those identifiers include the Network Address or ID of the VANET computer and the ID of the submission request tokens or ETokens. ETokens can be described as randomly generated unique serial numbers that are only known and generated locally at the VANET computer of a subscriber. Only x ETokens per subscriber are accepted by the data verifier per q length of submission intervals, *i.e.*, only 10-ETokens/second are accepted. The data verifier ideally should be a third party company that would operate as a contractor to a municipality. The parameter x is controlled by the network manager which establishes the x-times authentication of the system.

Table 3.1: Centralized Topology Variables

d:	info or data
s:	signing function - s signed, s' not signed
DV():	data verification function
DS():	data submission function
En():	encryption function
ID:	unique identifier of Etoken
UnID:	temporary unique identifier for Vanet Computer
t:	timestamp of information instance
loc:	location of information instance
q:	length of submission interval
x:	max # of request per interval q
m:	# of verification requested sent by vehicle in interval q
E(IDn):	global etoken database
M(UnID):	global blacklisted etoken machine database
ver	Global variable tracking # of verified data units
tok	Global variable tracking # of spent Etokens
I(d):	Data collectors private database of information

Once a piloted vehicle or non-piloted vehicle requests to send information to a verifier, an EToken is spent. That data verification request includes the following: d (data), t (timestamp of data instance), loc (location of data instance), ID (eToken ID), $UnID$ (VANET Computer eToken

Machine ID). Table 3.1 implies that En() is the encryption function signifying the content of the data packet is hidden from the receiver. The signing function is a Boolean s and s (s-not). A data verification request is represented as follows:

$$DV(En(s'(d), t, loc, ID, UnID) \tag{3.3}$$

Once the verifiers have received packet DV, as shown in Figure 3.2, the verifier checks if $unID$ has sent more than x data verification request in q time units. If condition is true, d remains unsigned and $unID$ is black listed for that submission interval. If false, the verifier continues to check if $unID$ is blacklisted by another verifier in the global verifier database. If true, d in DV remains unsigned and keeps s'. If false, the verifier continues by checking if the ID of the EToken ID is already in the global verifier EToken database. This step is done to prevent the reuse of ETokens. If ID is already in the database, that means that data has already been verified using that token by some verifier and d will keep s'. On the contrary if ID is not in the database, the verifier will add ID to the global database and sign d changing s' to s. The verifier sends the encrypted data back to $unID$. The piloted vehicle or non-piloted vehicle now has DV with $s(d)$. To enable the anonymous authentication the EToken Machine ID $unID$ changes in the next submission interval $q + 1$. Now the pseudo identifiers that the verifier had in previous intervals are useless since you cant cross-reference a randomly generated ID with an dynamic $unID$ that is only valid for a single submission interval. The vehicle to verifier process is implemented in Algorithm 3

Now to facilitate the anonymous data collection, data submission to data collector $DS()$ can be represented as follows:

$$DS(s(d(q-1)), t(q-1), loc(q-1), unID) \tag{3.4}$$

Data is submitted in the next interval q. That way the VANET Computers EToken machine will have a new $unID$ hindering the verifier and the collector from conspiring with their data to identifying subscribers in a network. With that said, an entity can perform the role of data verifier

Algorithm 3 Vehicle to Verifier

```
1: procedure INPUT(En(s'(d)), t, ID, UnID)
2:     if (m > n)
3:         UnID ← "No more request at this time"
4:     elseif (UnID == UnIDn)
5:         UnID ← "Request denied"                    ▷ Machine blacklisted
6:     elseif (ID == IDn)
7:         UnID ← "Duplicate request"                 ▷ Reused Etoken detected
8:         M().add(UnID)                              ▷ Add malicious ID to database
9:     else
10:        E().add(ID)
11:        En(s'(d)) = En(s(d))
12:        ver + +
13:        UnID ← En(s(d))
```

and collector without compromising subscriber data. Equation 3.4 shows that for data sent to the collector at interval q is the information from a previous interval $q - l$.

As shown in Figure 3.3, the data collector checks to see if the number of signed messages from the verifier and numbers of ETokens spent are equal to prevent verifier misconduct with their signatures. This condition is constantly checked by the data collector. Once the collector verifies the data is signed and valid, it stores the information to its database to distribute to subscribers or to sell to third parties. The vehicle to verifier is implemented in Algorithm 4.

Algorithm 4 Vehicle to Data Collector/Distributor

```
1: procedure (tok = E().size())
2:     INPUT: DS(s(d), UnID)
3:         while(tok ≤ ver)
4:             I().add(d)
```

This method of authenticating and signing with a data verifier then sending to a data collector helps both subscribers and data collectors by allowing them to receive important data from data collectors for improved VANET implementation and the data collector by allowing them to collect and possibly sell mass amounts of organic subscriber data while the identities of subscribers are safe.

3.2.2 Attack Prevention and Evaluation

To evaluate the effectiveness of centralized module of the hybrid topology methods were employed suggested in [1] such as attacker strength. For this thesis, attacker strength is expressed by:

$$s = \frac{x}{q} \cdot \frac{(A \cdot u)}{V} \tag{3.5}$$

To emulate real VANET scenarios, VANET Simulation tool and data collected from [1] was used. An area with an average vehicle density of 100 making $V = 100$ was surveyed. The traffic event being tested is a traffic jam and [1] suggest it take 10 seconds for the VANET algorithms to recognize a traffic jam making $u = 10$. As for x and q, the network operator controls those values, so for this test I chose $x = 10$ and $q = 1$. This implies 10 data requests can be sent per 1 second or millisecond. Attacker strength increases linearly with the increase in number of attackers. Due to this relationship, in a 100 vehicle scenario if $A = 0, 10, 20, , 100$ respectively, attacker strength would vary from 1 to 100. From this, one can deduce that attackers can only increase attacker strength by increasing the number of attackers. This is due to the x-times authentication scheme employed in the topology. Attacker effectiveness could further be diminished by the network operator by modifying the values of x and q. This displays the feasibility of hindering the submission of bogus data or brute force attacks in the centralized portion of the topology. Suspected high attacker density zones (*i.e.* $A = 100\%$) attackers strength can be weakened further by applying (3.2) to (3.5) yielding the modified attacker strength:

$$s2 = s \cdot R(t) \tag{3.6}$$

Figure 3.10 shows attacker strength at standard relevance $n = 1$. The decay of relevance is the same as any other surveyed neutral zone in the network.

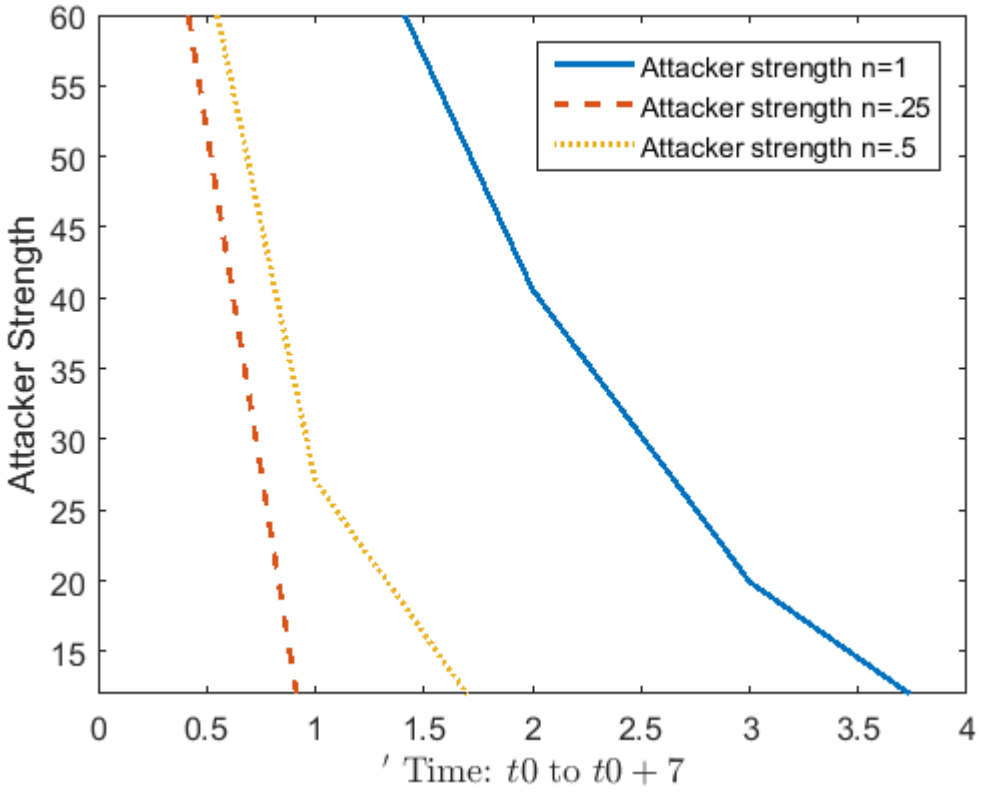


Figure 3.10: Attacker strength when unchecked and modified by network operator. The network operator can easily reduce attacker strength modifying the variable n

3.3 Results & Observations

Figure 3.10 shows attacker strength when the network operator modifies the n parameter in Equation (3.2) to $n = .25$ and $n = .5$. When $n = 1$ it is assumed that the network operator has not modified relevance. In Figure 3.10 ,you can observe a much swifter decay in relevance compared to the n parameter being left at 1. This displays the the network operators ability to address suspected high density attacker zones. Hence, it is less likely that the information from a suspected high-density attacker zone will skew the data in the overall VANET. This is another tool the network manager can use to ensure the validity of data on the network in the centralized portion of the topology. Data must meet certain relevance thresholds in order to continue to be forwarded

through the network Chapter 4 describes the impact of relevance in more detail.

Chapter 4

Framework for Secure Anonymous Data Transfer in VANETs

4.1 Motivation and Objectives

Reference [6] proposes an information centric network topology that dynamically determines which cars will be suitable data hubs based on popularity of information or consumer satisfaction. This proposal relies heavily on name specific data request coming from users in real time. Reference [6] defines information importance using Interest Satisfaction Frequency and Information Timeliness. Interest Satisfaction depends on users requesting specific data and a specific vehicle being able to satisfy that data request. However, this network model assumes vehicles are using each other as ad hoc network nodes to retrieve information from the Internet. VANET equipped vehicles have the ability to observe and gather about information from their local environment using various vehicle sensors without the need of Internet. Using the proposed framework, vehicles will be able to send receive useful information for navigation and smart vehicle decision making despite lack of Internet access via shared data peer to peer. Vehicles in this network will receive a great deal of information and that information must be stored locally and either shared or discarded. Reference

[7] addresses the life cycle of data in a network and proposes effective ways to manage stored data. Reference [8] speaks of methods to evaluate performance of a data centric network. Those concepts were adapted to develop a means of evaluated the routing quality in peer to peer ad hoc communication. The parameters in the smart data routing are temporal relevance and spatial relevance:

$$R(t) = \frac{(t+\lambda)e^{\frac{-t}{\lambda}}}{\lambda}, \tag{4.1}$$

$$R(d) = \frac{(d+\lambda)e^{\frac{-d}{\lambda}}}{\lambda} \tag{4.2}$$

Temporal Relevance and Spatial Relevance are the parameters that allow the network operator to set the characteristics of data flow in the centralized and decentralized modules of the network. Their values range across $0 < R(x) < 1$. This parameter allows the network operator to set constraints regarding how long messages are processed for decision making as well as for the distance data is routed from its location at specific instance. For example, if the network operator sets the filter to trigger an event at the threshold of .75, the network operator can set commands like purge data or archive data when message relevance drops below the set threshold. The network operator would be able to set different attributes in different areas of the network to address the unique needs of a particular city or location. This is illustrated in Figure 4.1, where each zone has been labeled with 1 through 9 accordingly. Messages in Zone 1 will have a longer lifespan as well as be forwarded farther than other messages in the network in respected to the other 8 zones in Figure 4.1. Once a message instance from that zone's temporal relevance drops below 0.3 the network operator will have a purge or archive action occur. Once the spatial relevance drops below 0.2 the network operator will also set a message to be archived or purged as well. With that said the $R(t)$ and $R(d)$ vary, and they could vary for different reasons. $R(d)$ in Zone 6, 8, and 9 are high to prevent messages from being forwarded to far from that zone. However, Zone 6 and 8 have slightly weaker constraints due to when the audit of the attacker zone was done, it determined more attackers were in Zone 9. This flexibility is put in place because a one size fits all approach could not accurately address the needs and scenarios in a real life network. As more

data is collected on audited zones and setting parameters a standardized network setup can be suggested and implemented.

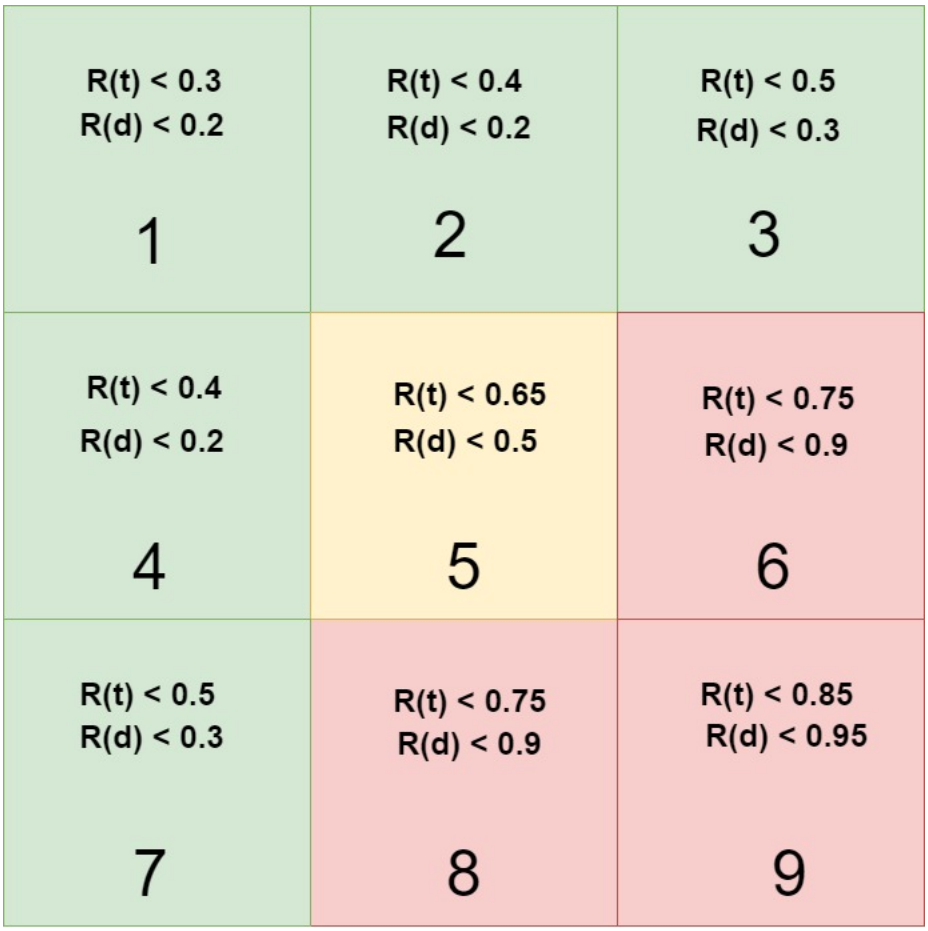


Figure 4.1: Example of network operator classifying zones in the network to modify the relevance of data.

The network operator can set the constraints on the life span of a message by adjusting the value of λ in Equation 4.1. For temporal relevance, the larger the λ value the longer the life of the information. For example, if t in $R(t)$ is observed in seconds and the network operator wanted to set the constraint of all messages stop forwarding from its time of instance in a surveyed zone after five seconds of its lifespan has passed. Assuming the default threshold is $if : R(x) < 0.5$ perform some action such as purge or archive. The network operator would simply have to set λ to 3. After 5 seconds have passed since a message's time of instance, R(t) for that message would roughly be less than 0.5. Once information passes the relevance threshold of $R(t)$ set by the

network operator, the applicable message can be purged from memory or archived and compressed in the vehicles VANET computer. This is done to optimize memory usage in a big data network and aim to ensure the network circulates the most useful and relevant information. The purge feature would be ideal in situations where a network operator wants vehicles to run mainly off of real time collected data. Real-time data collection would require a significant amount of data and if data storage optimization is not applied pro actively, it will be harder to scale to larger networks. An archive would be ideal if the network operator is trying to perform a statistical analysis on the behavior of the network. Collecting archived data from vehicles in the network can be valuable to assist in that endeavor. Subscribers can opt into archiving data for periods of time with some sort of incentive from the provider adding more symbiotic relationships between the subscriber and provider.

The network operator also has the ability to set constraints on the forwarding distance of information from its GPS instance in the V2V routing network by adjusting λ in (4.2). As with (4.1), the larger the λ the longer the forwarding distance. Note the that λ in (4.1) and (4.2) are not the same value and are independent of each other. With control over this parameter, the network operator can prioritize the forwarding distance of particular messages such as emergency messages to have a longer life span and have a longer forwarding distance to ensure that everyone in the city and network receive the important emergency broadcast. The assumptions of this framework are that vehicles have a V2V communication range of $+/-100$ meter radius [20] and they can store information locally in a VANET computer. In the next section, the proposed hybrid topology is going to be evaluated by employing evaluation concepts from [8]. The evaluation will survey routing quality which is a relation of VANET computer available memory to total memory capacity and data relevance and the amount of node hops the data has taken using (4.1) and (4.2). The decentralized nature of the VANET can lead to frequent packet loss so each time a packet is forwarded and broadcasted it has a chance of being lost in transmission. As a result, routing quality measures the effectiveness of message routing that uses a combination of the centralized and decentralized modules in the framework. Figure 3.4 shows how the centralized provider and the

V2V communication in the network work in unison to effectively route data to vehicle subscribers in the network.

4.2 Implementation & Methodology

For this evaluation, there will be three vehicle placement scenarios. Scenarios 1-3 are represented in Figures 4.2-4.4, respectively.

Figure 4.2: Scenario 1 for static implementation. Depicts vehicular scenario where vehicles are clustered and spaced out of the cluster

Figure 4.3: Scenario 2 for static implementation. Depicts vehicular situation were vehicles have close proximity to each other like a traffic jam

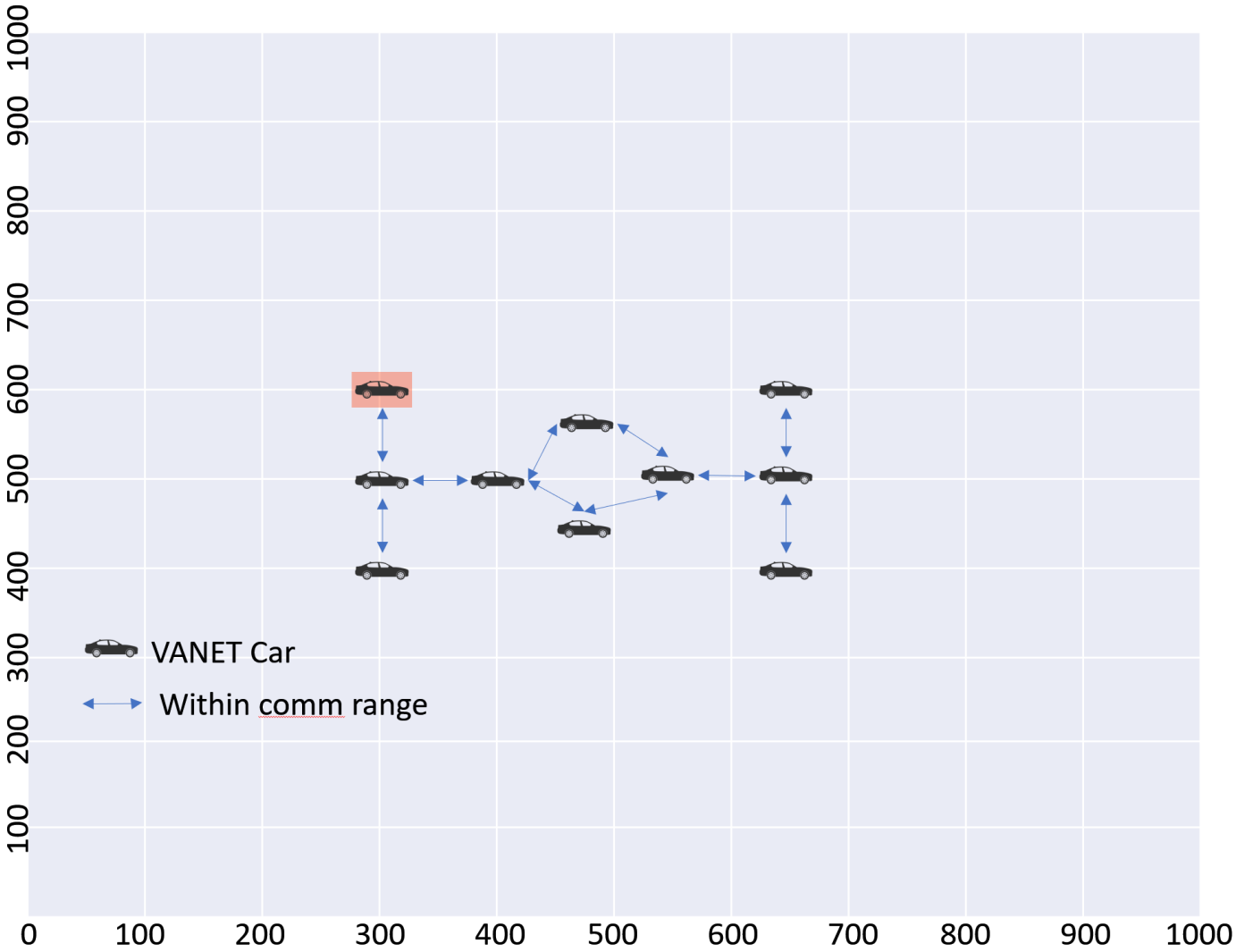


Figure 4.4: Scenario 3 for static implementation. Depicts vehicular scenario where vehicles are not closely cluster but are connected to all other vehicles V2V

It is assumed that data packets is worth 1 data unit and the VANET cars can only hold about 20 data units. The data units are stored locally at the vehicles and are purged or archived after their temporal relevance reaches the threshold value set by the network operator. It is assumed that purge and archive data frees up one data unit in this simulation. So for this simulation, the temporal relevance and spatial threshold will be set to 0.75, making the maximum life span for an information instance 0.6 time units. Therefore, the network operator will set the λ value to 0.624. A VANET car will stop forwarding a message after it has reached threshold value set by the network operator. For the simulation the maximum forwarding distance from location of instance is 200 distance units. The the λ value for $R(d)$ will be 208 so messages that reach a distance greater than 200 distance units will stop forwarding a message after it has reached threshold value set by the network operator.

The observed area for testing is 1000×1000 distance units and the test is observed for 2.5 time units in 0.1 steps. In scenarios 1,2,3 it is assumed that the 1000×1000 is a 100% attacker zone. The centralized data propagation of the hybrid topology has coverage area of the entire map. All cars are assumed subscribers of the data collector servicing the testing area. A V2V ad hoc broadcast message in this simulation is assumed to require 0.4-time units to be transmitted and received by neighboring nodes. The data collector/providers information requires 0.8-time units to be transmitted and received by all subscribers. The vehicle observed in this simulation is highlighted by the red square in figures 4.2-4.4. In Scenario 1, the observed vehicle is vehicle 4 from Table 4.1, in Scenario 2, vehicle 1 from Table 4.2 , and Scenario 3, vehicle 1 from Table 4.3.

The information packet from that vehicles observation will be forwarded via V2V to neighbors in range and eventually forwarded to all vehicles on the map using the data collector/provider. Each car has the ability to send messages. However, for the simplicity of the experiment, the routing quality with respect to information collected at the observed vehicle is being monitored. Routing quality is updated every time a new message is received. It is defined as the routing quality before the new message is received plus the routing quality of the new Q for each message received divided

Table 4.1: Scenario 1 Vehicle X,Y Locations in the 1000 × 1000 test area

Vehicle ID:	**X,Y Location:**
Vehicle 1	100,850
Vehicle 2	450,450
Vehicle 3	500,550
Vehicle 4	550,620
Vehicle 5	620,720
Vehicle 6	710,780
Vehicle 7	710,550
Vehicle 8	730,650
Vehicle 9	900,850
Vehicle 10	900,150

Table 4.2: Scenario 2 Vehicle X,Y Locations in the 1000 × 1000 test area

Vehicle ID:	**X,Y Location:**
Vehicle 1	150,550
Vehicle 2	200,550
Vehicle 3	250,550
Vehicle 4	300,550
Vehicle 5	350,550
Vehicle 6	450,550
Vehicle 7	500,550
Vehicle 8	550,550
Vehicle 9	600,550
Vehicle 10	650,550

Table 4.3: Scenario 1 Vehicle X,Y Locations in the 1000 × 1000 test area

Vehicle ID:	**X,Y Location:**
Vehicle 1	300,600
Vehicle 2	300,500
Vehicle 3	300,400
Vehicle 4	400,500
Vehicle 5	475,555
Vehicle 6	475,445
Vehicle 7	550,500
Vehicle 8	650,600
Vehicle 9	650,500
Vehicle 10	650,400

by the number of received messages. Q is defined in Equation (4.4) as the number of available data units A multiplied by a relevance of the message as it is received. This value is divided by the number of hops H or number of times the message is forwarded from its source. H is multiplied by 20 in Equation (4.4) to normalize the storage values therefore the maximum A for this experiment is 20. Routing quality can be expressed as the following:

$$RoutQual(N) = \frac{Q_N + Q_{N+1}}{N+1} \tag{4.3}$$

$$Q_N = \frac{A \cdot R\,(t)}{20 \cdot H} \tag{4.4}$$

4.3 Results & Observations

4.3.1 Static Simulation Results

The three vehicle scenarios depicted in Figure 4.2-4.4 were adopted from [4]-[6],[8] to appropriately evaluate frequent expected vehicular communication scenarios. This simulation has the parameters of Zone 8 and 6 in Figure 4.1.

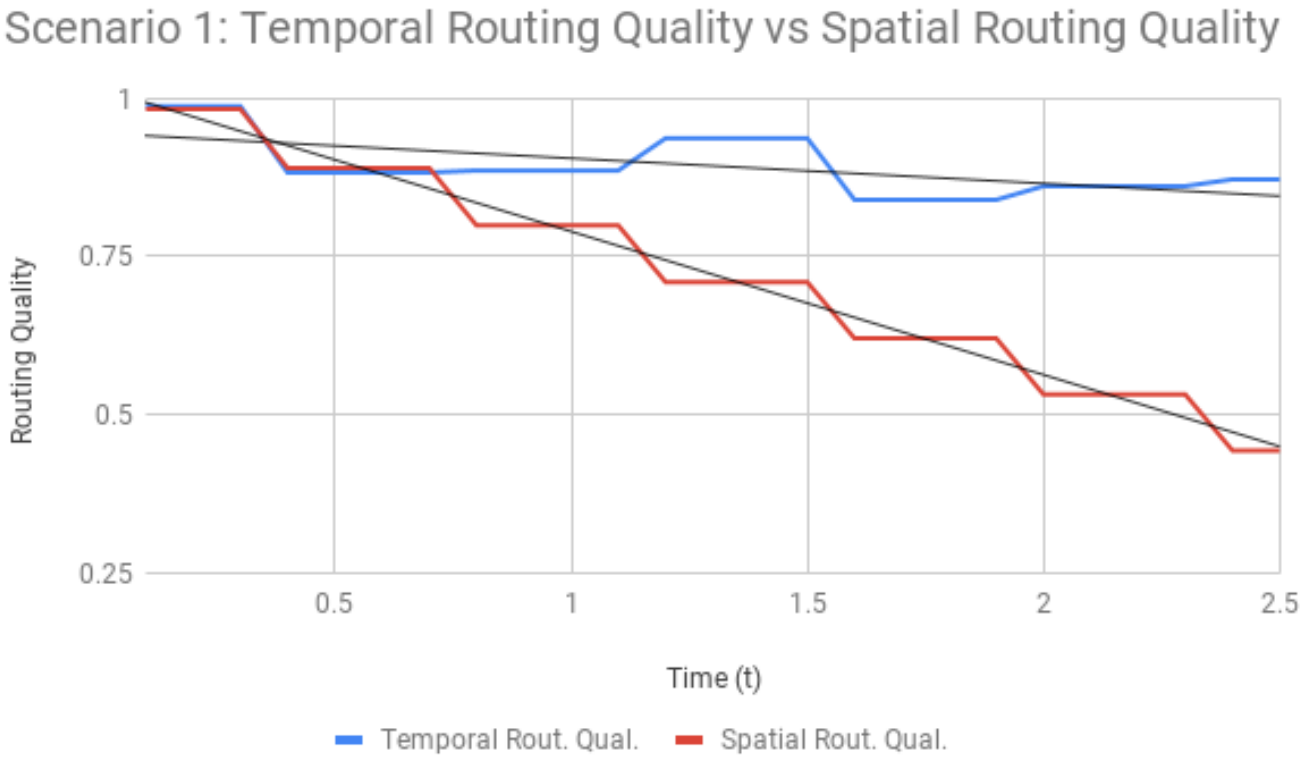


Figure 4.5: Static Simulation Result for Routing Quality in Scenario 1

In Scenario 1-3 it is assumed all the vehicles have a V2V communication range assumed to be a $+/-100$ meter radius [20]. In Scenario 1, the observed vehicle located at coordinate 550,620. The observed vehicle have 2 neighbors within VANET communication range. Each neighbor of the observed vehicle has at least one neighbor than can send data as well. Routing Quality in respect to temporal relevance performed well as the Routing Quality does not drop below 84% as

it fluctuates on a slight downward slope between the 80 and 90 percent. This shows that temporal relevance alone is a strong deterrent of malicious data in high attacker density zones. Though Spatial relevance performs well it still has a downward slope meaning that as time passes routing quality will continue to diminished until proper intervention is taken.

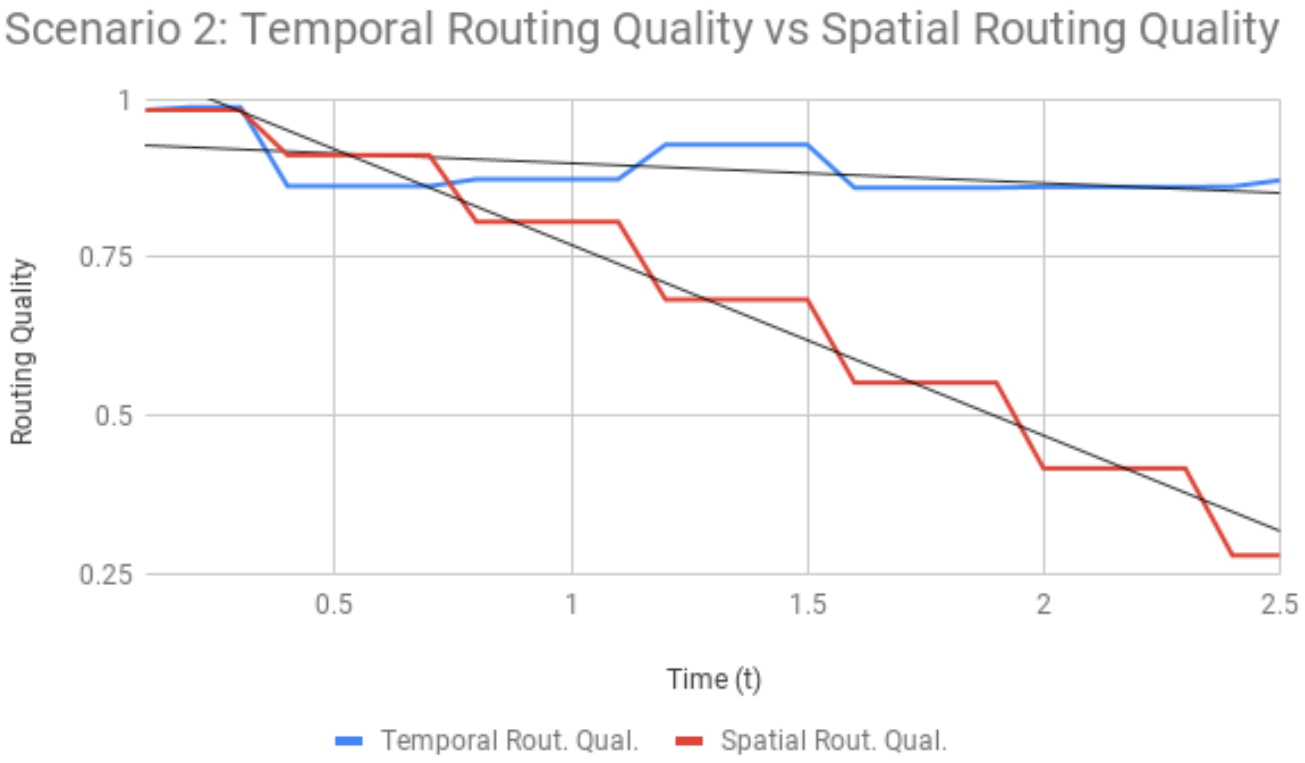


Figure 4.6: Static Simulation Result for Routing Quality in Scenario 2

Spatial Routing quality can make up for some of the shortcoming of the spatial routing. The distance constraint on messages can limit how far data leaves a zone. Therefore if a tight spatial constraint is set data will not be able to reach neutral data zones. Spatial relevance performs identically to temporal relevance until step 0.7. With spatial relevance, data is only purged or archived when the distance relevance threshold is crossed. This means that as long as your vehicle stays within the constraint range, data received from neighbors will not be removed or compressed. This can be especially problematic in traffic jam like scenarios as depicted in Figure 4.2. It is

apparent that spatial relevance performs the worst in scenario 2 due to more neighbors in proximity of the spatial constraint range. Spatial relevance is mainly strong for cleaning up whats old and useless and spatial relevance is good for compartmentalizing identified problem areas from neutral zone. When you apply both of these constraints on message forward you will have a robust method to address certain network problems.

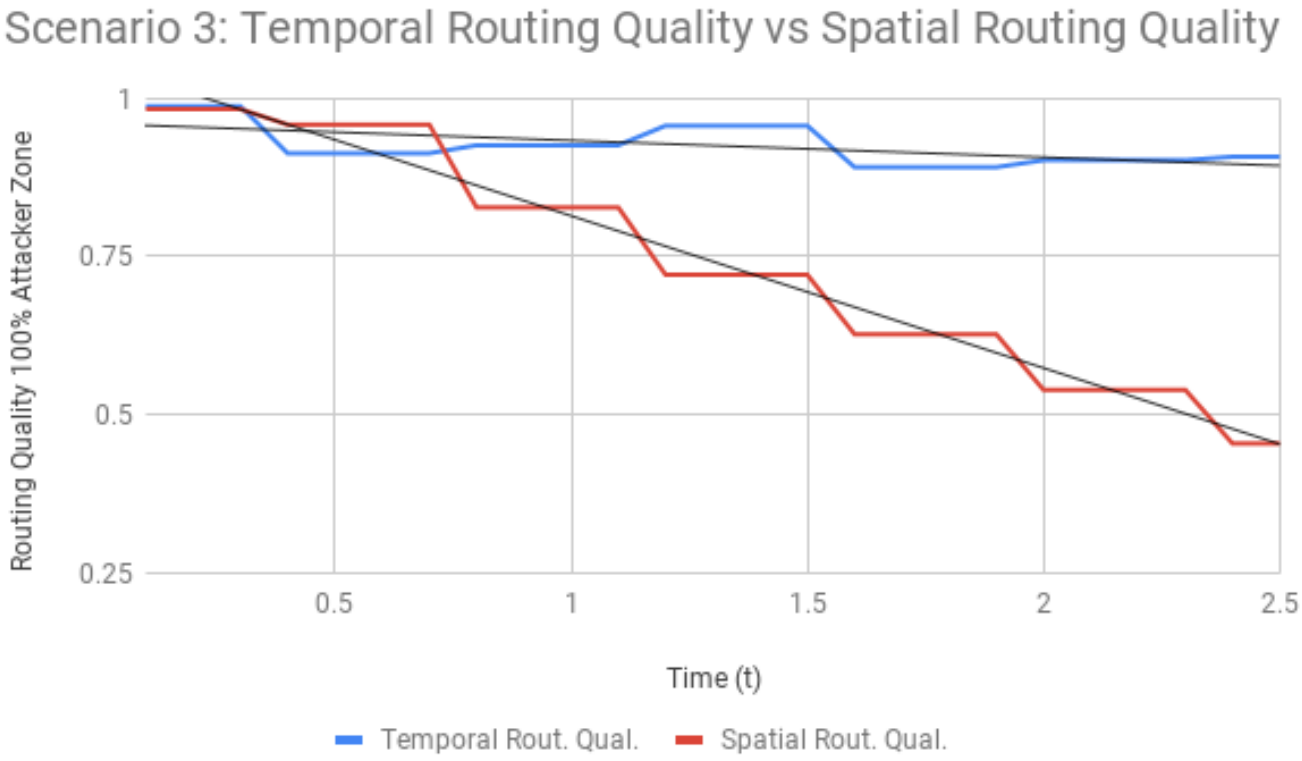


Figure 4.7: Static Simulation Result for Routing Quality in Scenario 1

Temporal Routing Quality performed well in all scenarios making it a good tool to control data flow and when a zone is being audited for possible malicious activity like the yellow zone shown in Figure 4.1. Until the zone has been determined to be red or high density attacker zone you don't want constrain the forwarding distance on possible good data. After a zone has been determine to be malicious spatial relevance can be applied to stop the fraudulent data from leaking to other neighbors in the VANET

4.4 Chapter Summary

VANETs are a practical solution to address safety and traffic optimization concerns. However, in order in implement a VANET a architecture must be put in place that protects user identity, network data integrity, and well as optimal network operation Network parameters such as relevance allow network operators to control data characteristics in a network. Network operators maintain network integrity by having data verified by data verifier in the network before a vehicle submits to the database. Data verifiers can use unique data verification request called Etokens to help maintain the integrity of the data submitted to the data collector. The Attacker strength simulation in this chapter displayed the ability of a network operator to mitigate areas in the network where known adversaries are located. The next chapter will go in more detail on how the network operator can address issues in the network.

Chapter 5

Conclusion & Future Work

5.1 Putting it all Together

This proposed VANET framework is a centralized/decentralized hybrid topology Vehicular Autonomous Network. As shown in Figure 5.2 ,vehicles collect data real time and forward it to it's nearest available neighbor. Figure 5.1 shows the though process of vehicles as they decided to send, receive, or process data. The spatial and temporal relevance are key parameters that allow the administrator of the network to control how data it routed and process in a respective network. As subscribers travel between city networks as shown in Figure 3.1, than VANET computer should be configured to only be able to participate in a network only if it is willing to accept that routing parameters set by the network operator of the network dwelt in. This permission can be accepted manually or dynamically. For a robust transition in foreign VANETs the permission granting process should be automatic. This is due to the parameters will cause no harm to the scriber because its is simply values such limit of data submissions per interval, relevance thresholds and so on. Since those parameters only serve to protect subscribers and support seamless VANET operation it behooves potential subscribers to automatically accept configurations while traveling.

Since all networks will not have high vehicle density which allows many nodes that do not have

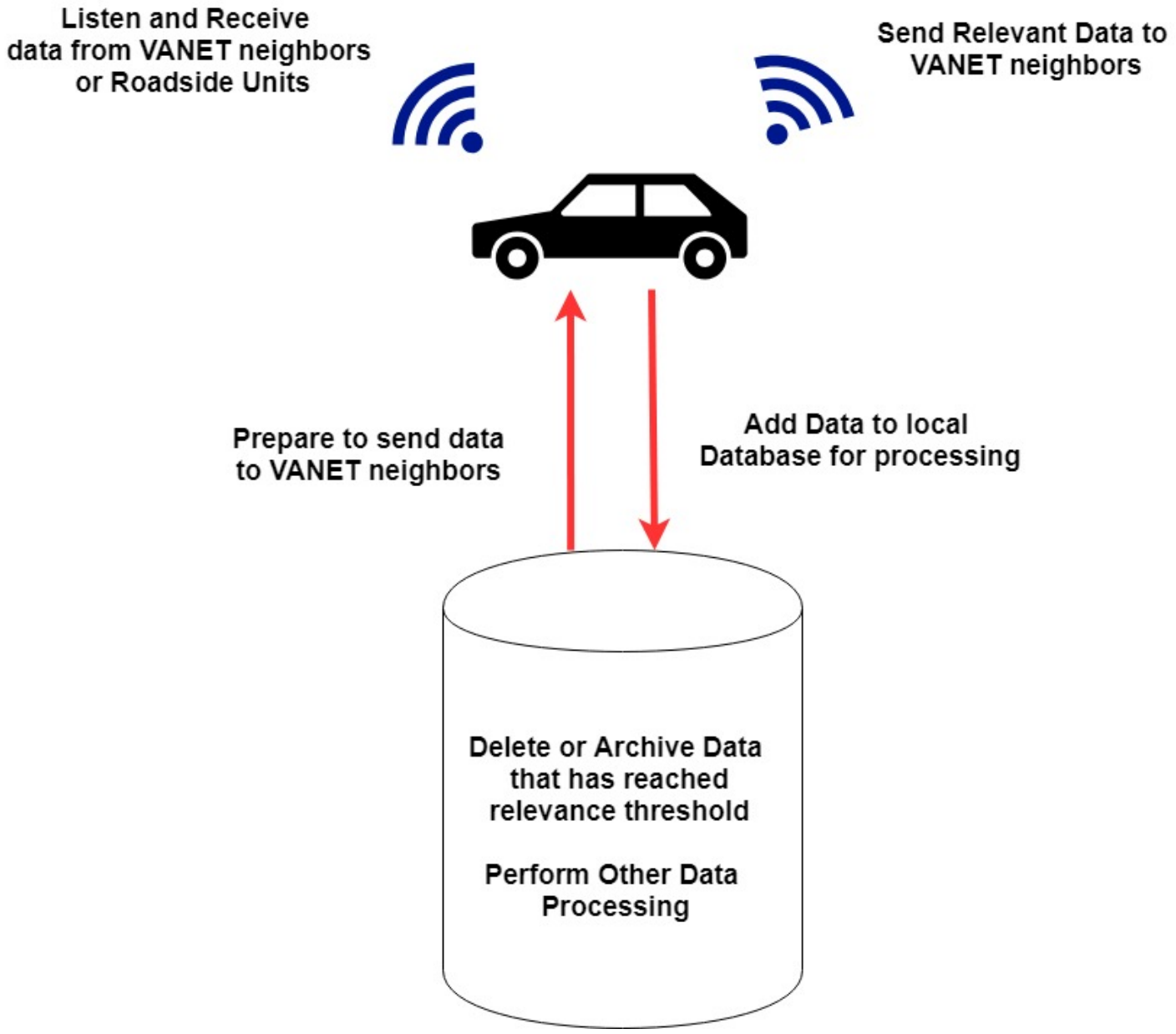


Figure 5.1: VANET Vehicle sending receiving and processing data.

a broadcast neighbor to receive useful data from the VANET as well as propagate data to other vehicles. The long range centralized routing can address the network availability issue. The municipality will deploy RSU units all over the range of the city providing overlapping network coverage as you can also see in Figure 5.2. Data sent through centralized means must be processed by the nearest data verifier for the municipality and then sent to the nearest data collector. Data collected goes to the centralized database, is processed then propagated throughout the network as seen fit. Messages verified must still follow the constraints of temporal and spatial relevance for more control over network characteristics. So if a message is collected at time $t = 0$, processed from $t = 0$ to $t = 0.8$ and the forwarding time constraint from time of instance in the zone instantiated is 0.6 the message will not be forwarded by the Road side units to other vehicles because it is set to be archived or deleted. The same goes for if the distance constraint from a message's location of instance is 2 miles, the data will not be forwarded to nodes longer than 2 miles away from the location of data instance. This help prevent data from the centralized municipal entity from being forward to parts of the network it is not needed. This further optimizes network resources.

Data collected from neighbor vehicles or RSU units are collected and processed locally at vehicles VANET computers. Autonomous vehicles will use collected data in conjunction with preinstalled self driving algorithms to make smart real time vehicle decisions. As the framework is applied and observed improvements can be made to the methodology. Success of this framework relies heavily on trust that most forwarded data will be useful for smart vehicle decisions. Therefore intervention methods applied in Chapter 4 are ways to combat possible aggressors in the network.

5.2 Summary of Thesis Achievements

My conference paper related to this thesis was published at IEEE VTC Chicago in the Fall Semester of 2108. It is a more premature point of view of the framework proposed in this Thesis. The Conference Paper is Titled "Secure Distributed Anonymous Data Collection for Vehicular Ad-Hoc

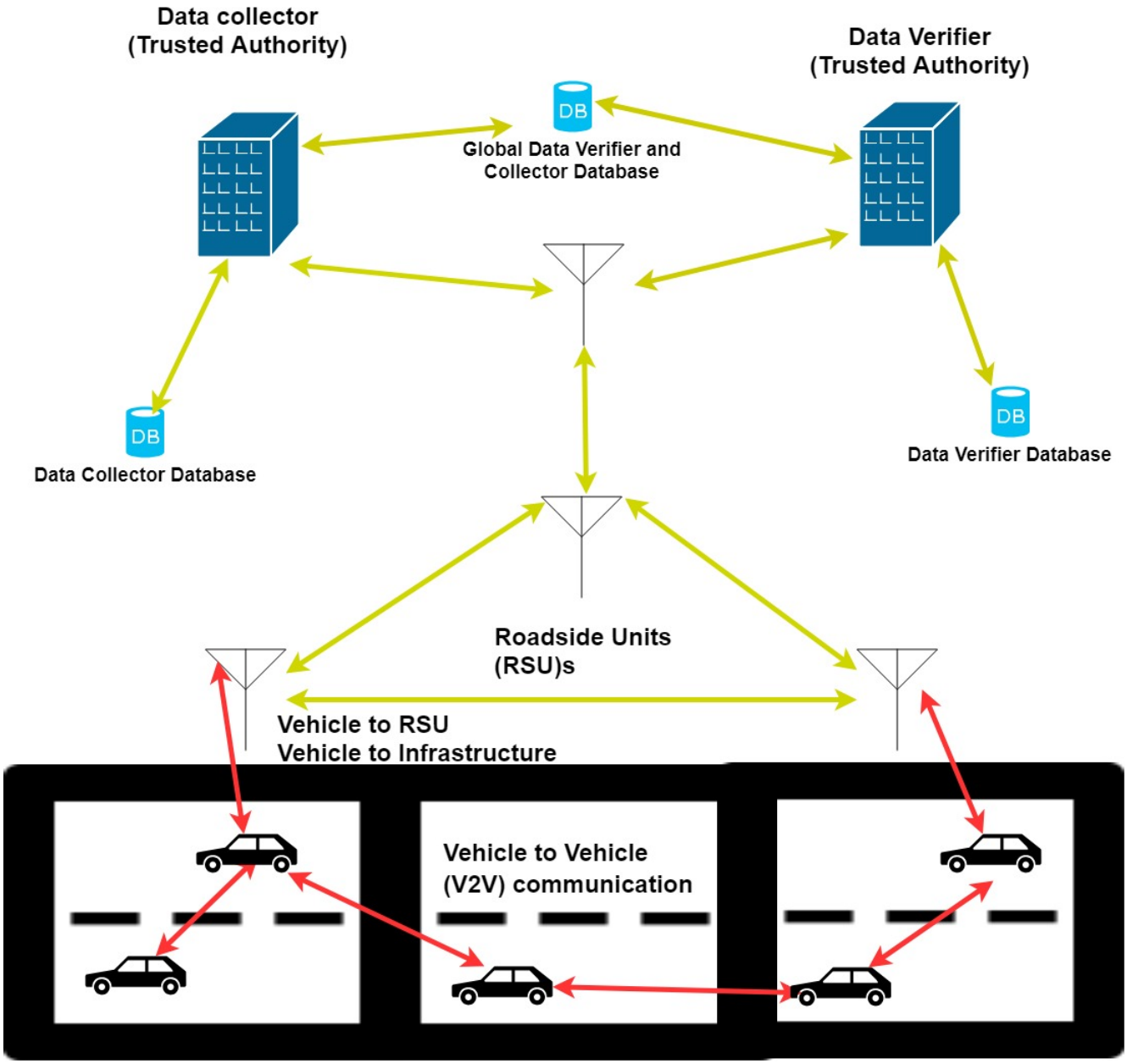


Figure 5.2: Depiction of Hybrid VANET Framework

Networks". It can be found search IEEE explore.

5.3 Concluding Remarks

Vehicular Communication is the catalyst to the future of Vehicular Autonomous Network Infrastructures. In order to promote more customer adaptation of this new framework small functions such as anonymity can give subscribers the piece of mind to accept the new technology. Just like any other computer network there will be malicious attackers so proper auditing and intervention methods must be put in place. When a municipality deploys this framework a direct attack on the network will be a punishable crime so after attacker zones are identified and culprits are apprehended it will slowly diminished the likely hood of malicious actors in the network overtime. Providing more reasons to support the network than attack the network will be key to the long term success of the framework. Subscriber incentives to perform task such as archiving data for later analysis or simply having your radio on to consent to being a node in a network can create a positive subscriber provider relationship

5.4 Future Research Direction

- We would like to evaluate more complicated vehicular scenarios to determine the true feasibility of applying this framework to a real test city. Therefore, we will perform the test applied in Chapters 3 and 4 in dynamic scenarios. We plan to develop a VANET simulator test bed to further evaluate the effectiveness of our proposed framework. We would like to develop a way to audit the network work to determine if there is malicious activity. TO do this we must also develop a better method to determine what data should be verified by a verifier or not.

- We would like to take the project in the direction of the automotive industry. Preferably

we would like to work with a automotive company to help with the development of V2V communication in the vision of the industry. A corporate sponsor would also afford us access to more resources to further develop and test the VANET framework. We also would like to go further in to the Cybersecurity required to protect the VANET based on minimum industry standards. Loss of client information can be a big negative to major companies and we must put everything in place in our framework to prevent that from happening.

- After a framework is fully defined, I would like to apply the framework to small controlled vehicle simulations and real life topology. Ideally we would like the initial deployment of this framework be done in a small controlled area then eventually to a small city. We have to take into account collisions, possible latency issues, and data merging issues from the centralized and decentralized parts of the hybrid network. Once the network model is fully developed it can be scaled and applied to larger VANETs.

Bibliography

[1] Andreas Tomandl, Dominik Herrmann, Hannes Federrath, PADAVAN: Privacy-Aware Data Accumulation for Vehicular Ad-hoc Networks 2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)

[2] AutoTalks. CRATON2 The Most Advanced Vehicle-to-Everything (V2X) Communication Solution. Autotalks Ltd., www.auto-talks.com/product/craton2/

[3] J. Camenisch, S. Hohenberger, M. Kohlweiss, A. Lysyanskaya, and M. Meyerovich, How to win the clonewars: efficient periodic n-times anonymous authentication, in Proceedings of the 13th ACM conference on Computer and Communications Security, 2006.

[4] Yu-Ting Yu, Mario Gerla, Information-Centric VANETs: A Study Of Content Routing Design Alternatives. 2016 International Conference on Computing, Networking and Communications, Mobile Computing and Vehicle Communications.

[5] Junaid Ahmed Khan, Yacine Ghamri-Doudane, Ali El Masri, Towards the Ranking of Important Smart Vehicles in VANETs - An Information-centric Approach, 2015 International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS).

[6] Junaid Ahmed Khan, Yacine Ghamri-Doudane, Dmitri Botvich. InfoRank: InformationCentric Autonomous Identification of Popular Smart Vehicles. IEEE VTC Fall 2015, Sep 2015, Boston,

MA, United States. Proceedings of IEEE 82nd Vehicular Technology Conference (VTC Fall), 2015, pp.6, 2015,

[7] Ioannis Psaras, Wei Koong Chai, George Pavlou, Probabilistic in-network caching for information-centric networks in ICN12 Proceedings of the second edition of the ICN workshop on Information-centric networking Pages 55-60

[8] [Grafling, Sebastian, Petri Mahonen, and Janne Riihijarvi. "Performance evaluation of IEEE 1609 WAVE and IEEE 802.11 p for vehicular communications." Ubiquitous and Future Networks (ICUFN), 2010 Second International Conference on. IEEE, 2010.

[9] D. Chaum, Blind signatures for untraceable payments, in Advances in cryptology. Springer, 1983.

[10] D. Chaum, Blind signature system, in Advances in cryptology. Springer, 1984, pp. 153153

[11] D. Chaum, Untraceable electronic mail, return addresses, and digital pseudonyms, Communications of the ACM, vol. 24, no. 2, pp. 8490, 1981.

[12] S. P. Borgatti, Centrality and network flow, Social networks, vol. 27, no. 1, pp. 5571, 2005

[13] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, Spatial tessellations:concepts and applications of Voronoi diagrams. John Wiley & Sons,2009, vol. 501

[14] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, Web caching and zipf-like distributions: Evidence and implications, in INFOCOM99, vol. 1. IEEE, 1999, pp. 126134

[15] Ross, P. E. (2014, May 29). Driverless Cars: Optional by 2024, Mandatory by 2044. Retrieved from https://spectrum.ieee.org/transportation/advanced-cars/driverlesscars-optional-by-2024-mandatory-by-2044

[16] Bhoi, S. K., & Khilar, P. M. (2013). Vehicular communication: A survey (Masters thesis, National Institute of Technology, Rourkela, Odisha 769008, India). IET Networks.

[17] Yan, G., Wang, Y., Weigle, M., Olariu, S., Ibrahim, K.: Wehealth: a secure and privacy preserving ehealth using notice. Proc. Int. Conf. Wireless Access in Vehicular Environments (WAVE), 2008

[18] Zeadally, S., Hunt, R., Chen, Y.-S., Irwin, A., Hassan, A.: Vehicular ad hoc networks (VANETS): status, results, and challenge, Telecommun. Syst., 2010, 50, (4), pp. 217241

[19] Zhang, L., Gao, D., Zhao, W., Chao, H.-C.: A multilevel information fusion approach for road congestion detection in VANETs, Math. Comput. Model., 2013, 58, pp. 120

[20] Kenney, J.B.: Dedicated short-range communications (DSRC) standards in the United States. Proc. IEEE, July 2011, vol. 99, no 7, pp. 11621182

[21] Ho, K.-Y., Kang, P.-C., Hsu, C.-H., Lin, C.-H.: Implementation of WAVE/DSRC Devices for vehicular communications. Int. Symp. Computer Communication Control and Automation, May 2010, vol. 2

[22] Morgan, Y.L.: Notes on DSRC & WAVE standards suite: its architecture, design, and characteristics, Commun. Surv. Tutor., 2010, 12, (4), pp. 504518

[23] Li, F., Wang, Y.: Routing in vehicular ad hoc networks: a survey, IEEE Veh. Technol. Mag., 2007, 2, (2), pp. 1222

[24] Nagaraj, U., Kharat, M.U., Dhamal, P.: Study of various routing protocols in VANET, IJCST, 2011, 2, (4), pp. 4552

[25] Diffie,W.and Hellman,M.E.New directions in cryptography. IEEE Trans. Information Theory 1T-22, 6 (Nov. 1976),644-654.

[26] Merkle,R.C. Secure communications over insecure channels. Comm. ACM 21, 4 (Apt. 1978),294-299

[27] Shepherd, A. (1970, August 24). What is network topology? Retrieved from http://www.itpro.co.uk/network-internet/31778/what-is-network-topology

[28] Hansler, E., et al. Optimizing the Reliability in Centralized Computer Networks. IEEE Transactions on Communications, vol. 20, no. 3, 1972, pp. 640644., doi:10.1109/tcom.1972.1091160.

[29] Abram, J., and I. Rhodes. Some Shortest Path Algorithms with Decentralized Information and Communication Requirements. IEEE Transactions on Automatic Control, vol. 27, no. 3, 1982, pp. 570582., doi:10.1109/tac.1982.1102987.

[30] Pramanik, Aniket, et al. Decentralized Topology Management on Mobile Ad Hoc Networks. 2015 Global Conference on Communication Technologies (GCCT), 2015, doi:10.1109/gcct.2015.7342635.58

[31] Dabek, F., Cox, R., Kaashoek, F., & Morris, R. (2004). Vivaldi: A Decentralized Network Coordinate System. ACM SIGCOMM Computer Communication Review, 34(4), 15. doi:10.1145/1030194.1015471

[32] Guimer, R., Daz-Guilera, A., Vega-Redondo, F., Cabrales, A., & Arenas, A. (2002). Optimal Network Topologies for Local Search with Congestion. Physical Review Letters, 89(24). doi:10.1103/physrevlett.89.248701

[33] HLDI, I. (2018). General statistics. [online] IIHS. Available at: https://www.iihs.org/iihs/topics/t/general-statistics/fatalityfacts/state-by-state-overview [Accessed 18 Dec. 2018].