NETWORKING AND SECURITY SOLUTIONS FOR VANET INITIAL DEPLOYMENT STAGE

by

BABER ASLAM
M.S. University of Central Florida, 2009

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Electrical Engineering and Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2012

Major Professor: Cliff C. Zou

ABSTRACT

Vehicular ad hoc network (VANET) is a special case of mobile networks, where vehicles equipped with computing/communicating devices are the mobile wireless nodes. However, the movement pattern of these mobile wireless nodes is no more random, as in case of mobile networks, rather it is restricted to roads and streets. Vehicular networks have hybrid architecture; it is a combination of both infrastructure and infrastructure-less architectures. The direct vehicle to vehicle (V2V) communication is infrastructure-less or ad hoc in nature. Here the vehicles traveling within communication range of each other form an ad hoc network. Whereas the vehicle to infrastructure (V2I) communication has infrastructure architecture. The vehicles connect to access points deployed along the road. These access points are known as road side units (RSUs) and vehicles communicate with other vehicles/wired nodes through these RSUs. To provide various services to vehicles, RSUs are generally connected to other RSUs and to the Internet. The direct RSU to RSU communication is also referred as I2I communication.

The success of VANET depends on existence of pervasive roadside infrastructure and sufficient number of smart vehicles. Most of the VANET applications and services are based on either one or both of these requirements. A fully matured VANET will have pervasive roadside network and enough vehicle density to enable VANET applications. But, the initial deployment stage of VANET will be characterized by lack of pervasive roadside infrastructure and low market penetration of smart vehicles.

It will be economically infeasible to initially install a pervasive and fully networked roadside infrastructure resulting in failure of applications and services that depend on V2I or I2I communications. Further, low market penetration will result in insufficient number of smart vehicles to enable V2V communication resulting in failure of services and applications that depend on V2V communications. Non-availability of pervasive connectivity to certification authorities and dynamic locations of each vehicle will make the security solutions, based on some central certificate management authority, difficult and expensive. Non-availability of pervasive connectivity will also affect the backend connectivity of vehicles to Internet or rest of the world.

Due to economical considerations, the installation of roadside infrastructure will take a long time and will be incremental thus resulting in a heterogeneous infrastructure with non-consistent capabilities. Similarly, smart vehicles will also have varying degree of capabilities. This will result in failure of applications and services that have very strict requirements on V2I or V2V communications.

We have proposed several solutions to overcome the challenges that will be faced during initial deployment stage of VANET: 1) a VANET architecture that can provide services with limited number of heterogeneous roadside units and smart vehicles with varying capabilities, 2) a backend connectivity solution that provides connectivity between Internet and smart vehicles without requiring pervasive roadside infrastructure or large number of smart vehicles, 3) a security architecture that does not depend on pervasive roadside infrastructure or a fully connected V2V network and fulfills all the security requirements, and 4) optimization solutions for placement of

limited number of RSUs within a given area to provide best possible service to smart vehicles. The optimal placement solutions cover both environments: the urban areas and the highways.

I dedicate this dissertation work to my family. Particularly to my wife and children, for their understanding, patience and encouragement throughout my PhD. A special thanks to my parents for their continued support and prayers.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr Cliff C. Zou for his continued guidance, support and encouragement throughout my PhD program here in University of Central Florida.

I would also like to thank members of my dissertation committee, Dr Mustafa Bassiouni, Dr Damla Turgut and Dr Morgan Wang, for their valuable guidance and suggestions on my dissertation.

I would also like to thank University of Central Florida for providing me resources, facilities and environment without which I would not have been able to complete my PhD.

I would especially like to thank my family for their continued support and encouragement throughout my PhD program.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Overview: Vehicular Networks

Wireless Networks can have infrastructure or infrastructure-less architecture. In infrastructure wireless networks, wireless nodes communicate with other wireless/wired nodes through an access point (AP) and no direct communication takes place between the wireless nodes. Whereas, in infrastructure-less wireless networks, there is no AP and wireless nodes communicate with other wireless nodes directly. The infrastructure-less wireless network is also known as ad hoc network or mobile ad hoc network (MANET) since the wireless nodes are usually mobile. Vehicles when equipped with computing/communicating devices (or On-Board Units - OBUs) also become mobile wireless nodes (sometimes referred as "smart car(s)" or "smart vehicle(s)") and the network becomes Vehicular ad hoc network (VANET). Vehicular networks are a special case of mobile networks; the movement pattern of mobile wireless nodes is no more random rather it is restricted to roads and streets.

Vehicular networks have hybrid architecture; it is a combination of both infrastructure and infrastructure-less architectures. The direct vehicle to vehicle communication commonly referred as V2V communication is infrastructure-less or ad hoc in nature. Here the vehicles traveling within communication range of each other form an ad hoc network. Whereas the vehicle to infrastructure communication also referred as V2I communication has infrastructure architecture. The vehicles

connect to access points deployed along the road. These access points are known as road side units (RSUs) and vehicles communicate with other vehicles/wired nodes through these RSUs. To provide various services to vehicles, RSUs are generally connected to other RSUs and to the Internet (like a distribution system in IEEE 802.11 architecture [1]). The direct RSU to RSU communication is also referred as I2I communication. An example of a VANET is shown in Figure 1.1.

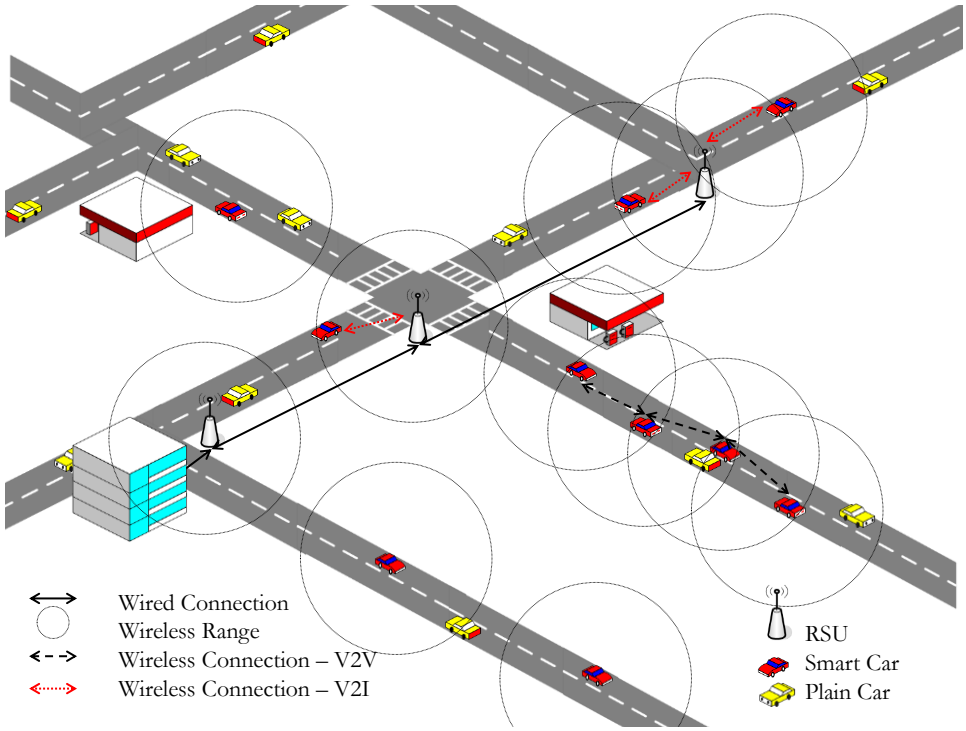


Figure 1.1: Vehicular ad hoc network - VANET

The architectural components of VANET include vehicles, roadside units (RSUs), I2I network, back-end connection of I2I network to Internet and some kind of management authority. The communication components of VANET include V2V communication, V2I or I2V communication and I2I communication. V2V and V2I communications complement each other. A pervasive RSU deployment with an elaborate I2I network will provide pervasive V2I communication that can

provide a service without any need of V2V communication. And on the other hand, a fully connected V2V network can overcome the absence of an elaborate V2I communication.

The nodes in VANET are highly mobile as compared to that of MANET but the mobility is restricted to roads/streets and is also affected by traffic conditions. The traffic conditions vary both temporarily and spatially. These characteristics result in highly dynamic network topology with frequent connections/disconnections. The presence of buildings and other manmade architectures alongside roads restrict the communication direction and range. The connection times are therefore very short due to high node mobility and existence of these obstacles. Vehicles tend to move in packets with large gaps between packets thus resulting in disconnected networks. As compared to MANET, VANET does not suffer from limitation of power and storage. As compared to other wireless/wired networks the applications and services of VANET are strongly related to location.

Efforts for standardization of vehicular communications/networks have long been done by different standard organizations such as International Standard Organization (ISO), European Telecommunications Standards Institute (ETSI) and Institute of Electrical and Electronic Engineers (IEEE). ISO's TC 204 WG 16 (Technical Committee 204 Working Group 16) is handling the standardization initiative and is called CALM (started as "Communications, Air-interface, Long and Medium range" and is now ""Communications Access for Land Mobiles") [2]. ETSI's TC ITS (Technical Committee on Intelligent Transportation Systems) has five Working Groups (WG1-5) working in this direction [3]. IEEE's 1609 WG (Dedicated Short Range Communication Working Group) is working to develop IEEE 1609 family of standards for vehicular networks also referred as WAVE (Wireless Access in Vehicular Environments) [4]. IEEE 1609 family of standard includes

IEEE 1609.0,1,2,3,4, IEEE 1609.11 and IEEE 802.11p [5-11]. IEEE 802.11p [11] amends IEEE 802.11 [12] for access in vehicular environments and is based on Dedicated Short Range Communications (DSRC) [12]. The relationship among WAVE Model and OSI Model is shown in Figure 1.2 [13].

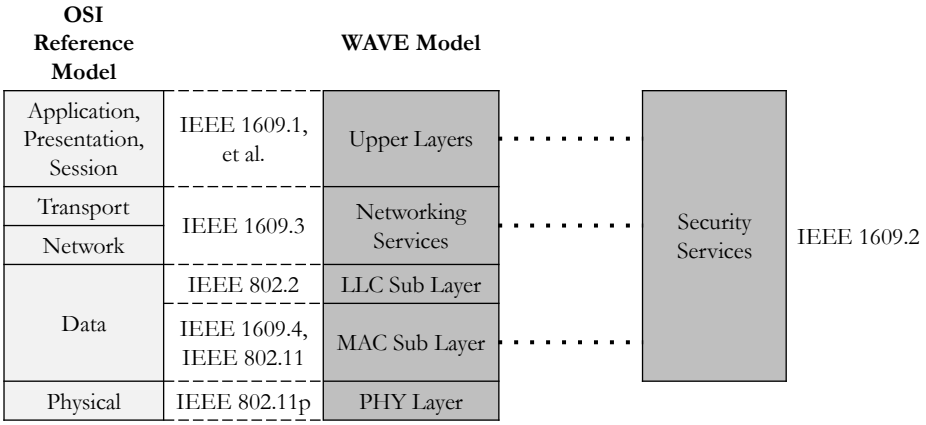


Figure 1.2: Relationship among WAVE Standards Family

1.2 Motivation

The success of VANET depends on existence of pervasive roadside infrastructure and sufficient number of smart vehicles. Most of the VANET applications and services are based on either one or both of these requirements. Initial deployment stage of VANET will be characterized by lack of pervasive roadside infrastructure and low market penetration of smart vehicles.

It will not be economically feasible to initially install a pervasive and fully networked roadside infrastructure. This will result in intermittent V2I and backend connectivity, causing failure of

services and applications that depend on this type of communication. The lack of infrastructure will remain a reality in rural areas and along highways even after the initial deployment stage. Further, backend connectivity will also remain an issue in rural areas and along highways even after the initial deployment stage of VANET.

Low market penetration will result in insufficient number of smart vehicles to enable V2V communication, thus causing failure of services and applications that depend on V2V communications (such as accidents report) [14]. Low density of smart vehicles will reduce the connection time between vehicles [15]. Low density of smart vehicles will also result in sparse and disconnected networks.

Expensive installation and maintenance of roadside infrastructure will make the services expensive to users, thus reducing the rate of market penetration that in turn will make investment in infrastructure further unattractive for service providers. There will be a need to optimally place the limited number of roadside units to provide best service to smart vehicles.

During the initial deployment stage, due to non-availability of pervasive connectivity to certification authorities and dynamic locations of each vehicle, it will be difficult and expensive to have security solutions based on some central certificate management authority.

Due to economical considerations the installation of roadside infrastructure will take a long time and will be incremental thus resulting in a heterogeneous infrastructure with non-consistent capabilities.

Similarly, smart vehicles will also have varying degree of capabilities. This will result in failure of applications and services that have very strict requirements on V2I or V2V communications.

1.3 Proposed Work and Contributions

The proposed work is to overcome the challenges that will be faced during initial deployment stage of VANET. The contributions of this dissertation are summarized below.

- Design of a VANET architecture that does not need expensive roadside infrastructure or large number of smart vehicles, can provide services with limited number of heterogeneous roadside units and smart vehicles with varying capabilities, is scalable and deployable.
- Provision of backend connectivity to Internet to smart vehicles without requiring pervasive roadside infrastructure or large number of smart vehicles, especially in rural areas and along highways.
- Design of security architecture that does not depend on pervasive roadside infrastructure or a fully connected V2V network. The architecture is economical, scalable and deployable.
- Optimal placement of limited number of RSUs within a given area to provide best possible service to smart vehicles. The optimal placement solution covers both environments: the urban areas and the highways.

1.4 Organization of Dissertation

Remainder of dissertation is organized as follows. Chapter 2 gives the design of an economical and deployable VANET architecture. Chapter 3 presents a backend connectivity solution using satellite

receive only terminals. Chapter 4 discusses security architecture for VANET initial deployment stage. Chapter 5 presents solutions for optimal placement of limited number of RSUs in two distinct environments; along highways and in urban areas. Finally, Chapter 6 gives the conclusion.

1.5 References

[1] IEEE STD 802.11-1997 IEEE Standard for Information Technology—Telecommunications and information exchange between systems-Local and metropolitan area networks - Specific requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 1997.

[2] Online: http://www.isotc204wg16.org/home - accessed September 2011.

[3] Online: http://www.etsi.org/website/Technologies/IntelligentTransportSystems.aspx - accessed September 2011.

[4] Online: http://standards.ieee.org/develop/wg/1609_WG.html - accessed September 2011.

[5] IEEE P1609.0 – Draft Standard for Wireless Access in Vehicular Environments (WAVE) – Architecture.

[6] IEEE 1609.1-2006 - IEEE Trial Use Standard for Wireless Access in Vehicular Environments (WAVE) - Resource Manager.

[7] IEEE 1609.2 -2006- IEEE Trial Use Standard for Wireless Access in Vehicular Environments (WAVE) - Security Services for Applications and Management Messages.

[8] IEEE 1609.3 -2010 – IEEE Standard for Wireless Access in Vehicular Environments (WAVE) - Networking Services.

[9] IEEE 1609.4 -2010- IEEE Standard for Wireless Access in Vehicular Environments (WAVE) - Multi-Channel Operations.

[10] IEEE 1609.11-2010 - IEEE Standard for Wireless Access in Vehicular Environments (WAVE)-- Over-the-Air Electronic Payment Data Exchange Protocol for Intelligent Transportation Systems (ITS).

[11] IEEE 802.11p – IEEE Draft Standard for Information Technology -Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 6: Wireless Access in Vehicular Environments.

[12] ASTM E2213-03 - Standard specification for telecommunications and information exchange between roadside and vehicle systems - 5GHz band dedicated short range communications (DSRC) medium access control (MAC) and physical layer (PHY), September 2003.

[13] IEEE P1609.0 D05 - Trial Use Standard for Wireless Access in Vehicular Environments (WAVE) – Architecture, February 2008.

[14] N. Wisitpongphan, O. Tonguz, F. Bai, P. Mudalige, and V. Sadekar, "On the Routing Problem in Disconnected Vehicular Networks," INFOCOM'07, May 2007.

[15] M. Mabiala, A. Busson, and V. Vèque, "Inside VANET: hybrid network dimensioning and routing protocol comparison", VTC'07-Spring,, Apr 2007.

CHAPTER 2 AN ECONOMICAL AND DEPLOYABLE VANET

Although there is plenty of research on VANET, the solutions to the challenges existing during the long transition period in VANET deployment are largely ignored. There are some research solutions that offset the absence of roadside architecture by either not relying on it (i.e., using only V2V communication), or using cellular architecture, or using existing available Wi-Fi hotspots, or using static/mobile relay units (Delay tolerant networks)[9, 10, 12, 16, 18, 24]. Most of these approaches target specific applications and may not be easily upgradable (during later stages) to VANET architecture (such as those defined by IEEE P1609 working group). There is also lack of research on how to provide incentive for vehicle owners to install wireless devices when support from roadside infrastructure is insufficient and very limited number of smart vehicles are on the road.

We present an economical and deployable VANET system design to solve these challenges. Our focus is to provide a transitional/interim solution that can be used to start up (or give impetus) to VANET activities during the long initial transition period by making VANET easy to deploy, secure and economical. At the same time the design should be incremental/progressive and should be easily transformed into architectures that are already specified in VANET standards/protocols without requiring any major revamp/modifications. From now on we will mostly refer "smart vehicles" as "vehicles" without considering any vehicle that has no communication device, unless there is a need to explicitly mention smart vehicles.

The contributions of this solution include are a VANET system design that is economical, realistic, incremental and deployable during the initial long transition period when smart vehicles have a low penetration rate. Core component in our design, the Roadside Units (RSUs), can be standalone with minimum intelligence in their basic form. Our proposed design does not require RSUs to be interconnected or connected to the Internet. We present a basic protocol that makes the communication between roadside units possible via mobile vehicles. The simulation results indicate considerable performance gains by just using standalone RSUs.

The chapter is organized as follows. In section 2.1 presents related work. Section 2.2 gives detailed description of our proposed design. Section 2.3 presents the simulation details. Section 2.4 presents discussion and section 2.5 gives the conclusion.

2.1 Related Works

Most of the existing research in VANET presents routing algorithms for V2V communication [2-4]; these protocols rely on the assumption that sufficient number of vehicles will be available for relaying messages. Some of the research work also addresses the routing in disconnected or intermittently connected networks [5-7]. A hybrid approach has also been presented to address limited connection time between vehicles [8]. However, during the initial deployment there will not be sufficient number of smart vehicles on the road to even form small clusters for these protocols to work. Further, lack of roadside infrastructure will also make use of hybrid protocols difficult.

The technique for transmission of data between nodes of a disconnected or partitioned network using temporary storage at intermediate nodes is a delay tolerant network (DTN) [9]. Besides satellite networks, the concept of DTN has been widely applied to VANET (which may be considered as DTN especially during initial deployment stages) [5, 9-13]. It is pertinent to highlight here that most of the existing research in this context is on V2V communication protocols [5, 10, 11] where mobile nodes temporarily store the message if no route is available and later opportunistically forward the message. These protocols may be used to solve the disconnected network problem due to uneven distribution of traffic, but may not be an effective solution to low penetration rate issues. Throwboxes in UMass' DieselNet [12] are similar to our standalone RSUs. The throwboxes act as stationary routers to improve connectivity among mobile nodes (buses) that are equipped with multiple radios (including a long-range radio), GPS recording devices etc. In our research the RSUs are not just the routers, but in addition they also receive, process and disseminate information (such as safety or warning). Mobile nodes in MIT's CarTel [13] are equipped with multiple sensors; the data collected from these sensors is processed and transferred to a central portal by these nodes. The transfer is accomplished opportunistically via Wi-Fi (hotspots, roadside units), Bluetooth or by nodes themselves (data mules). A specially designed delay-tolerant network stack (CafNet) is used for communication. Our emphasis is not on making major modification to existing VANET standards/protocols, but to enable their gradual/incremental deployment during initial phases.

Infostations architecture allows use of high speed and generally dispersed access points. The access points/stations afford transfer of high volume of data at cost of connectivity. They can be especially useful in VANET environment where vehicles are moving at fast speeds and connection time to

access points is very limited [14, 15]. This architecture cannot solve the low penetration problem since the infostations will generally be widely dispersed. Further, these must also be fully networked with Internet, which will be expensive to install and maintain.

A number of researches have incorporated cellular networks in VANET [16-19]. Cellular networks are mostly used as a backbone - a replacement to roadside infrastructure. Cellular networks, though pervasive, offer lower data rates as compared to Wi-Fi (roadside infrastructure). Although with the advent of 3G/4G technologies data rates close to broadband can be achieved, these technologies are not uniformly available throughout cellular coverage areas and many users are still dependent on other heterogeneous technologies (WAP, GPRS, EDGE, HSDPA, etc [20]). Further, cellular data plans subscriptions are expensive; an unconstrained plan with a 5GB/month limit costs approximately $700/year. 5GB/month means per day a user on average can send/receive 50 emails (20 with attachments), download a song and a game/app, view 40 web pages, posts 10 social media posts with photos, and watch a streaming video of 40 minutes [21]. Although unlimited data plans and those that cost few ten of dollars are also available, but these have several fine print conditions; such as 'usage patterns' (no file sharing, excessive usage, etc), 'can only be used on smart phones' (no tethering), 'can only access certain service' (email, predefined websites, etc), and 'must have a qualifying voice plan'. Some service providers are charging approximately $2/MB or 1¢/KB for web browsing. All major cellular service providers are now offering and encouraging users (such as offering 'unlimited Wi-Fi usage with data plan') to access data through hotspots (which use Wi-Fi just like VANET roadside infrastructure instead of 3G/4G). This also highlights cost/benefit of Wi-Fi over 3G/4G. In addition cellular networks also have few other disadvantages such as expensive to built/maintain, billing/licensing issues among different service providers, higher roaming rates,

large and variable latency, central switching/resource management, difficult to scale and occasional blackouts [20, 22-27].

A class of protocols uses the store-and-forward approach for V2V communications [2, 28]. MDDV [2] is a multi-hop V2V protocol. It uses predictability of vehicle movement to route the messages. It assumes the vehicles to be equipped with GPS and digital maps. It uses trajectory and geographical based forwarding. If end-to-end path does not exist then messages are stored and later forwarded when a connection is established. VADD [28] is also a multi-hop protocol using the carry-and-forward paradigm. In this protocol a vehicle carries a message until it finds another vehicle in communication range, then it forwards the message. It assumes that the vehicles are equipped with GPS, digital maps and also have detailed traffic statistics such as vehicle density, vehicle speeds. It bases its decision of message-forwarding on these statistics. Both the protocols [2, 28] are used to transfer messages between vehicles in multi hops.

Lochert et al. [29] compare the performance of standalone and networked stationary supporting units (SSU) in context of low penetration rate. The work focuses on dissemination of information from a central point in a city scenario. They show that the networked SSUs (connected via a backbone) improve the performance dramatically as opposed to the standalone SSUs. V2V communication also plays an important part in their scenario. Whereas in our case we used very limited penetration rate so that V2V communication is not possible and our results show that standalone RSUs do increase the performance.

Our work comes closer to protocols that use vehicles to transfer messages between roadside units [30-33]. M.C. Chuah et al. [32] present a protocol using multi-hop V2V communication between roadside units. They present a detailed mechanism for forwarding of messages at each hop. It makes use of query and response messages at each hop. B. Petit et al. [30] present a set of protocols for data relaying between roadside units using vehicles. The protocols give different options for transfer of data between a source/sink and a vehicle. It uses solicit and beacons for selection of appropriate vehicle to carry the data. The work has been further extended by A. Mansey et al. [33] giving vehicle-roadside unit data transfer mechanisms and reliable multi-packet data transfer schemes. The protocol does not provide details on routing between different roadside units. Y. Ding et al. [31] present a static node assisted adaptive routing protocol. It is basically a multi hop protocol that makes use of static nodes at the intersection to store and forward the messages, thus improving performance over other multi hop V2V communication protocols.

Our research work differs from above mentioned protocols in many ways. We do not assume vehicles to be equipped with GPS and digital map, or have road statistical data, which makes our design more realistic especially in the initial transition period. Our design does not involve V2V communication, thus it works well when smart vehicles are sparsely distributed on roads. We do not assume roadside units to be always connected to infrastructure (i.e., fully networked or connected to the Internet), which makes the RSU deployment in our design economical and practical during the transition stage. We present an integrated design involving vehicles and roadside units with varying degree of capabilities. Besides being economical, the design is also scalable and can easily be upgraded without any major modifications in protocol.

2.2 Proposed Design

The common characteristic of all VANET applications is either collection or dissemination of information from/to vehicles in a timely and efficient manner. V2V and V2I communications complement each other in achieving this flow of information. For example, we can overcome the issue of low market penetration of smart vehicles by having more elaborate roadside infrastructure (i.e., passing information through fully networked roadside units or using infrastructure to infrastructure – I2I communication), or conversely, high penetration can overcome lack of roadside infrastructure or I2I communication (i.e., passing information using V2V communication). As discussed earlier, during the initial stage of VANET, both V2I (also I2I) and V2V communications will not be very effective. So we need to address the issues of V2V and V2I connectivity in an efficient and economical way. Since we cannot influence the market penetration of smart vehicles, the other solution is to improve V2I and I2I communications; which will, in turn, complement the lack of V2V communication. One option is to have pervasive fully networked roadside infrastructure (to improve V2I and I2I communications). Though it may be possible to have such a network in urban-areas but in rural-areas/along-highways (where there is not much manmade infrastructure) this option will be quite expensive and impractical. Another option is to use cellular network as a replacement to roadside infrastructure. Though cellular networks are pervasive, but in addition to the technical and economical disadvantages mentioned earlier in section 2, this option will also introduce heterogeneous technologies (i.e., typical VANET architecture in urban-areas and cellular based architecture in rural-areas/along-highways); making the transition to final VANET architecture difficult.

We suggest improving connectivity/communication by using roadside units (RSUs). In its basic form our proposed RSU is standalone with only store-and-forward capability, which makes it economical and easy to install/maintain, especially in rural-areas or along highways. Other types of RSUs include those that are locally connected to each other or connected to the Internet (those located close to manmade infrastructure or in urban-areas). Details of RSU design are given in Section 4.1. An overview of RSU's role in achieving different connectivity requirements is given below:

- **V2V Communication**: Direct multi-hop V2V communication will not be possible during initial deployment phases due to low market penetration. V2V communication among vehicles with spatial displacement will be improved using store-and-forward capability of RSUs and those with temporal displacement will be improved if RSUs are networked. Besides broadcast communication, one-to-one communication among temporally displaced vehicles may also be achieved with the support from RSUs if vehicles have fixed routes and travel schedules (as in the case of daily commute).
- **V2I Communication**: Economical and easy installation/maintenance features of RSUs help in achieving high RSU densities even in rural areas and along highways; this will help in improving V2I communication.
- **I2I Communication**: V2I communication is of little use if there is no I2I communication. RSUs that are connected to each other or to the Internet can easily communicate with each other. Standalone RSUs can use passing-by vehicles as relays to communicate with each other.

Our design integrates RSUs of varying capabilities thus making architecture economical, easy to install/maintain, incremental/progressive (basic RSUs can easily be upgraded to higher capability ones), homogenous (same technology in urban and rural areas) and upgradable to final VANET architecture.

2.2.1 Roadside Unit - RSU

The motivation of our design is to make roadside units light weight, simple/easy to install and economical. Our proposed design does not require all RSUs to have the same capabilities. Multiple versions of RSUs enable engineers to have necessary flexibility in designing a VANET architecture that is suitable to their requirements and budget.

2.2.1.1 Multiple Versions

We define several different versions of RSUs with varying capabilities/functions and network connectivity. In its basic version, an RSU is a standalone unit with temporary store-and-forward capability. In terms of connectivity, RSUs can be standalone, locally networked (via wire or wireless such as WiMax [44]), connected to the Internet via wire or wireless, or just have backend receiver-only capability in order to receive data from satellite, cellular, commercial radio, etc. RSUs may have sensors for monitoring local weather, road condition, traffic, etc. All RSUs are tamper proof, capable of receiving and sending data from/to vehicles and have some information processing capabilities. Possible versions of RSUs are listed in Table 2.1. A possible architecture with standalone, locally connected and globally connected RSUs is shown in Figure 2.1.

Table 2.1 Different version of RSUs with increasing functionality. An RSU with a larger version number will be more expensive but provide more functionality.

Version	Store and Forward-Repeater	Intelligent with information processing	Sensors to collect local data- traffic, met etc	Limited local connectivity	Backend Receive only- Radio, Satellite	Backend duplex connectivity
1.0	Yes	No	No	No	No	No
1.1	Yes	No	No	Yes	No	No
1.2	Yes	No	No	Yes	Yes	No
2.0	Yes	Yes	No	No	No	No
2.1	Yes	Yes	No	Yes	No	No
2.2	Yes	Yes	Yes	Yes	Yes	No
3.0	Yes	Yes	Yes	Yes	Yes	No
3.1	Yes	Yes	Yes	Yes	-	Yes

Store-and-forward is the basic capability and enables an RSU to transfer messages between spatially and temporally displaced vehicles. Intelligent information processing gives RSUs the capability of encryption/decryption, data verification, provision of time/location stamp, certificate revocation, etc. Sensors are used to collect local traffic and weather data. This collected data can be used for verification of data provided by vehicles. Limited local connectivity means an RSU is connected to at least one adjacent RSU. "Backend receive only" enables reception of critical safety information, certification revocation lists, etc. It is an economical way to receive important non-local messages for dissemination to vehicles in an area, such as fire, flood, and earthquake emergency warnings. It can also be used for distribution of certificate revocation lists to RSUs similar to [34]. "Backend duplex connectivity" means connection to the Internet; such RSU can send and receive data

to/from the Internet. Other RSUs can connect to the Internet through the backend duplex connected RSUs.

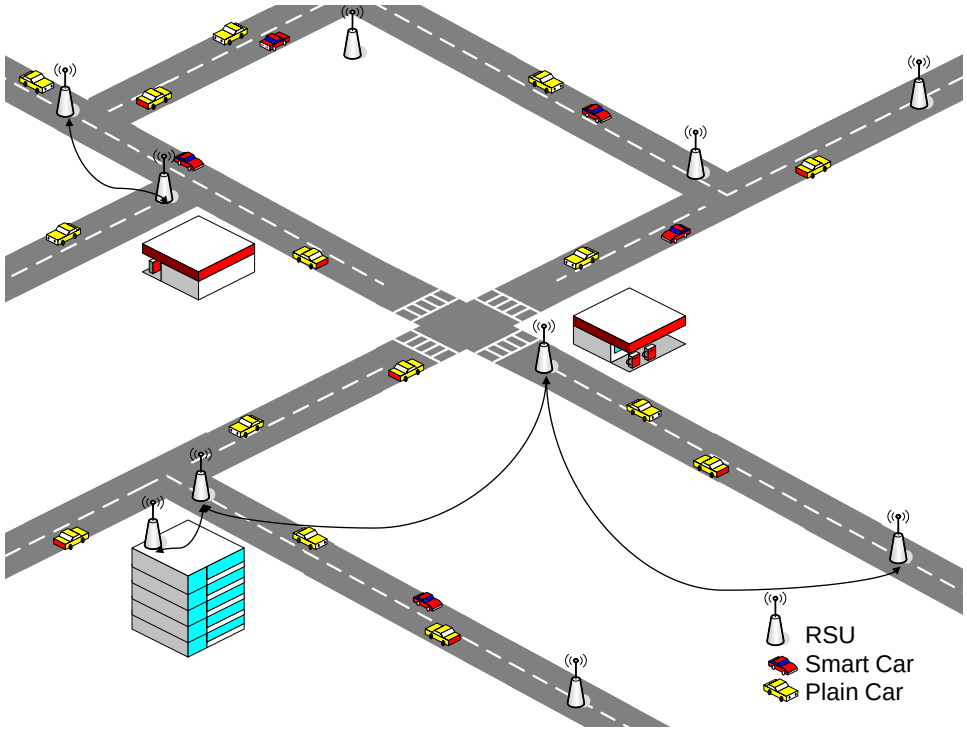


Figure 2.1: The proposed architecture consists of RSUs deployed along the roads. RSUs can be standalone (the three RSUs on top right), locally connected to adjacent RSU (two on the up-left corner), or connected to Internet infrastructure (three on the bottom). What versions of RSUs to install depends on overall budget and services we want to provide.

2.2.1.2 Deployment

Different versions of RSUs help in achieving economical deployment across diverse areas. In urban areas where Internet connectivity is pervasive, it is economical and easy to deploy Internet connected RSUs, whereas in rural areas or along highways where it is difficult/expensive to extend Internet connectivity (though a limited number of RSUs may be connected to Internet through

cellular network), it is more economical to deploy standalone or locally networked RSUs. These basic standalone RSUs will help in getting VANET started and later they may be replaced with more advanced ones without any system overhaul.

Each RSU will have a distinct identification and an associated digital signature certificate. The certificate issuing authorities for RSU's certificates may be organized on area/region basis. All vehicles in VANET will have certificate-issuing-authority's certificate and will be capable of verifying RSUs' certificates and messages signed by these certificates. At startup, a VANET can even have Version 1.0 RSUs without any certificates. These RSUs can be used for only store-and-forward functions and be deployed at non-critical locations.

Each RSU will be aware of local map, its own location and locations of other RSUs in the area. Additionally, each RSU will also maintain a routing table with known path to each of the other RSUs in the area. Initially, this information will be added at the time of installation and later it will be updated periodically via RSU update messages. For this purpose, each RSU will periodically exchange signed Hello messages (containing routing table etc) with its neighbors. Routing-table-update procedures from any existing table-driven routing protocol may be employed for routing table updates and the details are hence omitted here. Routing between standalone RSUs relies on the relay by passing-by vehicles and is directly related to traffic density. If traffic density varies considerably during different times of day then the routing table may contain multiple entries accordingly (e.g., one each for morning commutes, one for evening commutes, and one for rest of the day).

2.2.2 V2V Communication

V2V communication is an important part of VANET; many VANET applications (such as cooperative driving, and safety warnings) depend on V2V communications. However, there will be a very small number of smart vehicles during the initial deployment phase of VANET and V2V communication will hardly exist. In such case standalone RSUs (with store-and-forward capability) can play an important role in achieving limited V2V communication. A sending vehicle sends a message to a nearby RSU, which stores and later forwards it to another passing-by recipient vehicle. Though this type of communication cannot be used for time-critical messages but it can still serve as a means to broadcast non-time-critical messages. If a vehicle has its certificate then it may also sign the message to ensure its authenticity to a receiver. In this way, a malicious vehicle transferring fake messages will be held accountable.

If an RSU is networked (locally or with Internet), then the RSU can support V2V communication between spatially displaced vehicles. This could be useful in quick dissemination of information within the network or across networks (if the RSU is connected to the Internet).

2.2.3 V2I (Vehicle to RSU) Communication

Our proposed design enables service providers to deploy a relatively large number of RSUs with less investment thus enabling more V2I communications. Each RSU will advertise its existence and offered services by broadcasting periodical beacons. The services offered by a particular RSU will depend on its version/capabilities, e.g., an Internet connected RSU may offer email service whereas

a standalone RSU may only offer store-and-forward service. If an RSU is capable of sensing nearby vehicles, it can broadcast its beacon only when a vehicle enters its broadcast range — this conserves power in low traffic conditions. The beacon broadcasting interval (*BI*) can be defined by the maximum allowed driving speed (*s*) and broadcast zone diameter (Z_d), i.e., $BI = Z_d / s$.

The beacon will include an RSU's ID, certificate, location, current time, location of adjacent RSUs, services offered and critical safety information. Critical safety information is included in beacon to reduce the information relaying time. The beacon message will be signed by its issuing RSU. Critical safety information messages may also be broadcasted independent of the beacons. In this case, critical safety messages will be given priority over other messages. They will be signed by sending RSUs and will include location of sending RSUs and current time. Vehicles may relay these messages to other passing-by vehicles.

2.2.4 I2I (RSU to RSU) Communication

I2I communication plays a vital role in both V2V and V2I communications. I2I communication may be considered a part of V2I communication especially when roadside infrastructure is fully networked. We consider I2I communication separately because in the initial deployment stage of VANET RSUs are not necessarily connected to each other or to the Internet. Data transmission will normally be limited to adjacent RSUs only. However, there may be situations when a message is needed to be sent to another RSU that is many hops away, such as sending information about a malicious vehicle to an RSU that is known to be connected to the Internet, or relaying an accident report to emergency vehicle that is known to be located near a particular RSU.

I2I communication, depending on the connectivity of RSUs, can be divided into two types, i.e., I2I Direct communication and I2I Indirect communication, which will be introduced next.

2.2.4.1 I2I (RSU to RSU) Direct Communication

Some RSUs may be locally connected to adjacent ones. This connectivity can be wired or wireless. Local connectivity is economical as compared to global connectivity (to the Internet). If two RSUs are connected to each other then direct I2I communication will be used. For this, existing protocols (such as those defined by IEEE P1609 working group) may be used, and the details are hence omitted. There may be a case where part of networking route is connected and part is disconnected; the connected part will use direct communication whereas the disconnected part will use indirect communication introduced in the following.

2.2.4.2 RSU to RSU Indirect Communication (via Vehicles)

If RSUs are not locally connected, then the RSUs communicate with each other using passing-by vehicles. A reputation system may be used to solicit vehicles' cooperation in relaying these messages. The exact details of such a reputation system are out of the scope and are not discussed in this paper. The addressing information will include the destination RSU's ID and its location. If the message is not for adjacent RSUs but several hops away, routing information will also be included. Routing information will include locations and IDs of all intermediate RSUs along the path to the destination RSU. The message will be signed by its originator and any confidential information will be encrypted. The certificate of the originator will also be appended with the message.

The basic idea of opportunistic routing is used in our design. In opportunistic routing as opposed to deterministic routing, the node that forwards a message is not predetermined. It is determined on the fly, normally by a subset of nodes that receive the broadcast [45, 46]. An RSU broadcasts a message to every vehicle in range. There are two possible options for the selection of relaying vehicle. In the first option, after receiving the message, each vehicle waits for a random amount of time and then acknowledges the message. On hearing the acknowledgement sent by one vehicle all other vehicles will discard the message. Therefore, only one vehicle that acknowledges first is selected as the message-relay-vehicle.

One possible problem can occur for this option when the relaying vehicle diverts from the route before delivering the message. In this case the probability of success can be increased by letting more than one vehicle to relay the message. Another possible issue is the hidden-node problem (note that the small number of smart vehicles during the initial stage of VANET deployment will reduce the chances of having a hidden-node); in this case more than one vehicle will acknowledge and carry the message. This operation will provide redundancy to the protocol, but at the same time, it will require duplicate suppression at the destination. (Mathematical analysis of the number of nodes required to deliver a message with certain probability of confidence is discussed later).

Acknowledgement messages will be restricted to only one hop. End-to-end acknowledgement may be included as an optional service. The calculation of acknowledgement timeout is discussed later.

2.2.4.2.1 Operation

When an RSU broadcasts a message, each receiving vehicle compares the destination location with its direction of travel and discards the message if it is for an RSU on the opposite direction. For a vehicle not equipped with GPS, we have two options to determine its direction of travel relative to the location of destination RSU. First, the vehicle can use the location of RSU it has just passed and the location of current RSU to determine its direction of travel. Second, in the message the sending-RSU can include the previous-RSU's ID that a vehicle must have passed, if it is along the desired direction.

A relaying vehicle passes the message to each intermediate RSU that is listed in the routing path of the message. When an RSU receives a message, it checks message integrity and then sends an acknowledgement to its immediate upstream RSU according to the routing information contained in the message. If the message has been received before, it is discarded and only the acknowledgement is sent. This ensures duplicate elimination on a per hop basis.

If the message receiving RSU is not the destination RSU, it rebroadcasts the message to the next RSU in the routing path. It then waits for an acknowledgment from the next RSU; waiting time is defined by acknowledgement-wait-time (details in next section). If no acknowledgement is received till the expiration of acknowledgement-wait-time, it rebroadcasts the message. The process is then repeated for a fixed number of times. This guards against network overloading since there may be the cases when a message has been received but acknowledgement cannot be sent due to lack of upstream vehicular traffic. The acknowledgement generated by the destination RSU may be sent

back to the source RSU as an optional service. An example flow of message and its acknowledgements is shown in Figures 2.2 and 2.3.

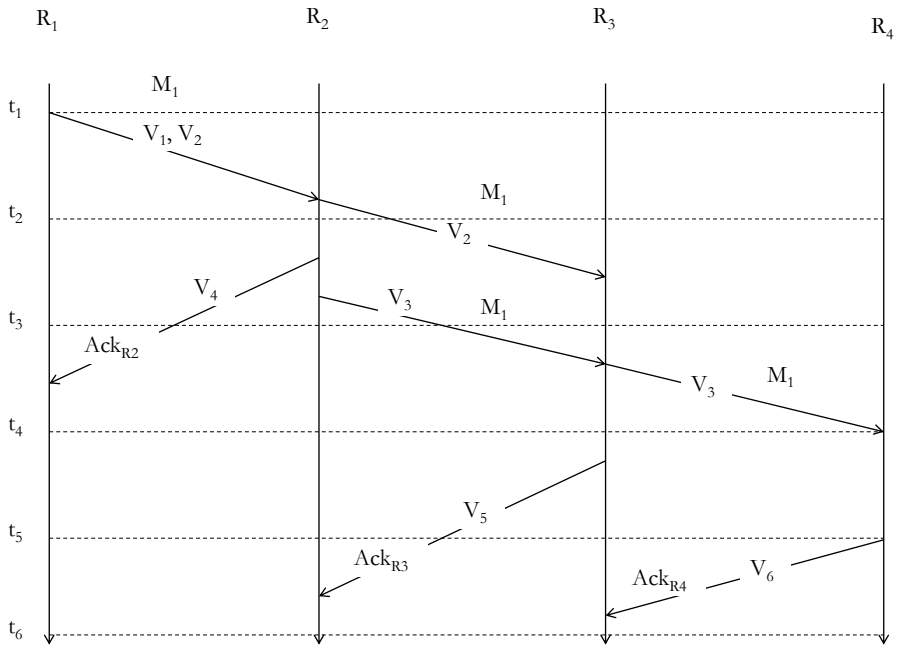
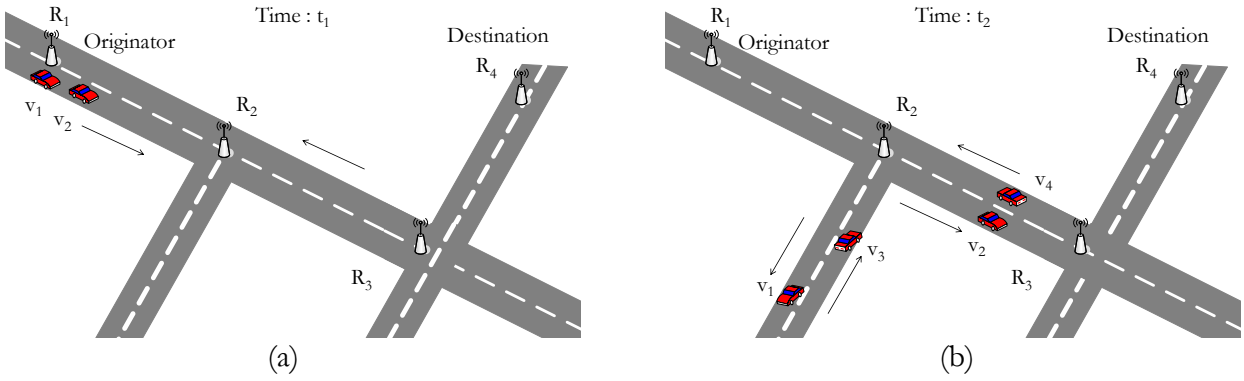


Figure 2.2: Flow of a message and its acknowledgement between RSU_1 and RSU_4 (illustrated also in Figure 2.3). Message M_1 is sent from RSU_1 to RSU_4 via RSU_2 and RSU_3. t_i represents time in order. R_i represents RSU_i. *Vi* represents vehicle *i* that that carrys a message. Ack_{Ri} is the acknowledgement message from RSU_i to RSU_{i-1}.



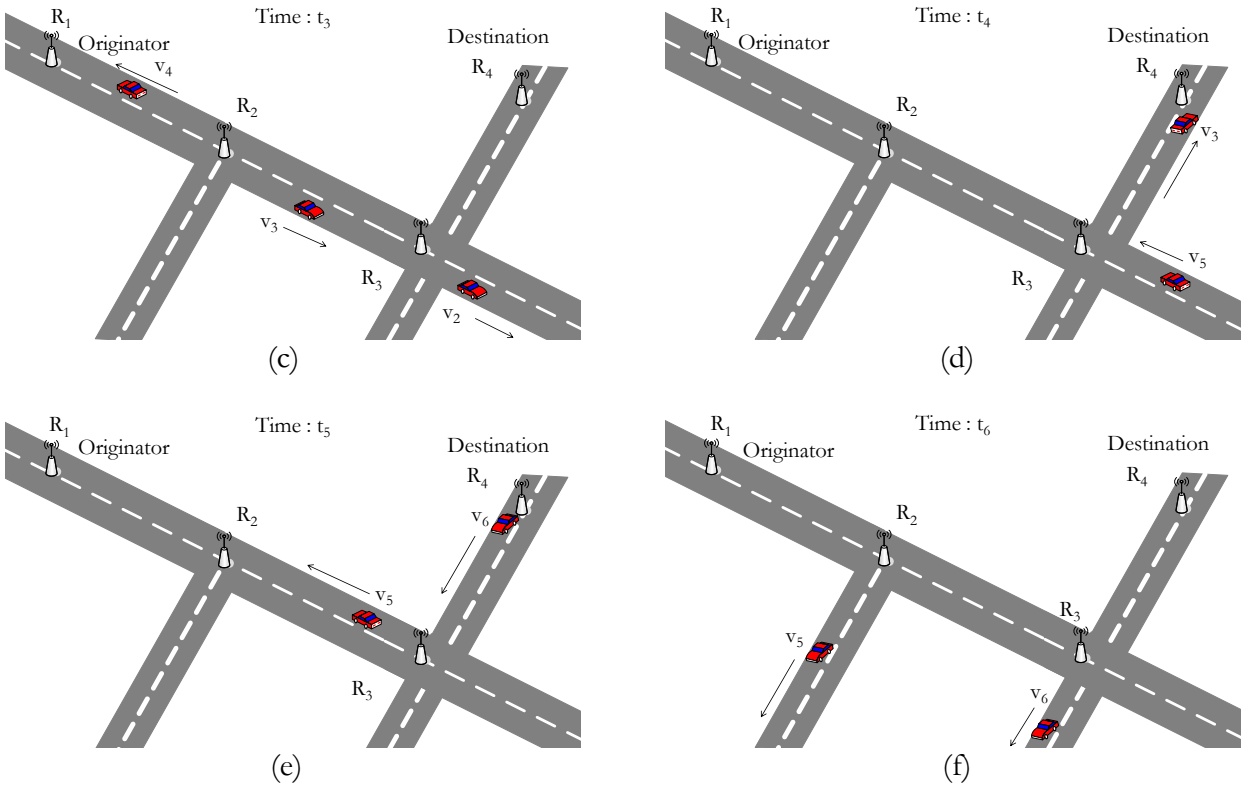


Figure 2.3: Snapshots of the road conditions when a message is sent from RSU_1 to RSU_4 via RSU_2 and RSU_3 (illustrated also in Figure 2). (a) V_1 and V_2 receive the message from RSU_1. (b) V_1 and V_2 deliver the message to RSU_2, V_1 diverts to its right at road junction, V_3 and V_4 approach RSU_2. (c) V_2 delivers the message to RSU_3, V_3 receives the message from RSU_2, V_4 carries the acknowledgement message from RSU_2 for RSU_1. (d) V_3 delivers the message to RSU_4, V_5 approaches RSU_3. (e) V_6 receives the acknowledgement message from RSU_4 for RSU_3; V_5 carries the acknowledgement message from RSU_3 for RSU_2. (f) The acknowledgement messages delivered by V_6 and V_5 to RSU_3 and RSU_2 respectively.

A vehicle may deliver the same message to more than one consecutive RSUs, for example, in Figure 2.2 and 2.3, vehicle V2 delivers message M1 to RSU2 and RSU3. In order to take advantage of this situation, each receiving RSU waits for acknowledgement from its next RSU on the routing path before re-broadcasting the message, since the vehicle that has delivered the message may also deliver the message to the next RSU. But if the traffic density is low, the receiving RSU may rebroadcast the message before the end of the wait timer (to simplify the logic the RSU may rebroadcast the message before starting the wait timer).

2.2.4.2.2 Acknowledgement Wait Time

Each RSU waits for its acknowledgement before retransmitting. The wait time (Wt) depends on the distance to the next RSU, the average speed of vehicles and traffic conditions. It is directly related to distance (L) and inversely related to vehicle speed (s) and traffic density (d) (upstream),

$$W_t = 2\frac{L}{s} + \frac{1}{ds} + \varepsilon \tag{2.1}$$

where ε is a constant which caters for processing done at a node before sending the acknowledgement.

The final wait time will be estimated using equation (3). Here α is the smoothing factor, M is the acknowledgement arrival time and D is the smoothed deviation (similar to TCP round-trip-time estimation model [47])

$$D = \alpha D + (1-\alpha)|W_t - M| \tag{2.2}$$

$$\text{TimeOut} = W_t + 4 \times D \tag{2.3}$$

2.2.4.2.3 Number of Relay Vehicles

We cannot be sure that a vehicle which has passed the source RSU and is carrying the message will always pass the destination RSU without diverting on the way. Therefore, we want to estimate the number of times a source should relay the message to have some degree of confidence that the message will reach the destination.

Suppose between two RSUs, there are one or several road diversions. Among the traffic flow entering from the source RSU, only *p* fraction of flow goes to the destination RSU. *N* represents the number of vehicles passing the source RSU; the random variable *X* represents the number of vehicles that have passed by the source also pass the destination RSU (Figure 4.4). Let's find out how many vehicles (*N*) should the source RSU ask to carry the message, in order to let the destination RSU have at least k vehicles passing through it, with a confidence of probability Pc (such as 95%).

Because each vehicle has an independent probability *p* to go to the destination RSU, the random variable *X* follows Binomial distribution. If we denote $f(n;N,p)$ as the probability of exactly n vehicles going through the destination, then according to Binomial distribution, we can derive:

$$f(n;N,p)=\binom{N}{n}p^n(1-p)^{N-n} \tag{2.4}$$

The question we asked above means that the probability of having less than *k* vehicles passing through the destination RSU must be no more than $1-P_c$. Thus the following inequality formula must be satisfied:

$$f(0;N,p)+f(1;N,p)+\ldots+f(k-1;N,p)\le 1-P_c \tag{2.5}$$

For $k>1$ (which will be the case if we want more than one vehicle to deliver the message for redundancy or security purposes) Equation 2.5 does not have a closed-form solution. To derive the value of *N*, we can test $N=1$, $N=2$, $N=3$,..., until we find the smallest value of *N* satisfying the formula.

When $k=1$, the above formula means that the value of N must satisfy:

$$(1-p)^N \le 1-P_c \tag{2.6}$$

$$or \quad N \ge \frac{\log(1-P_c)}{\log(1-p)} \tag{2.7}$$

Equation 2.7 gives the minimum number of vehicles that required to carry a message in order for at least one vehicle passing the destination with certain confidence level. Figure 2.5 shows the number of vehicles required for different values of p.

2.2.4.2.4 Protocol Simplification based on GPS Data

The large scale use of GPS technology has made GPS devices economical; it is likely that in the near future all modern vehicles will be equipped GPS devices, which provide valuable data such as up-to-date location, direction and speed. In addition, when a GPS device is used for navigation, it can provide destination and trajectory information. The additional data can be used to make the communication protocol simpler and more efficient.

An RSU can query a passing-by vehicle for destination and trajectory information. Based on this information, the RSU can decide whether or not to choose the vehicle for forwarding messages to other RSUs. This will reduce the number of vehicles used to relay any given message. One possible implication of this is privacy; the owner of a vehicle may not want to disclose the vehicle's destination or trajectory information to RSUs. This can be easily resolved as follows: When an RSU offers a message to a vehicle, it also includes the destination information of the message. The vehicle

can then reply either "YES" if it can carry the message or "NO" if it cannot carry the message based on its driving trajectory. The vehicle may also choose to reply "Do not Know" if it does not want to disclose any information about its destination; in this case, the original protocol can be used.

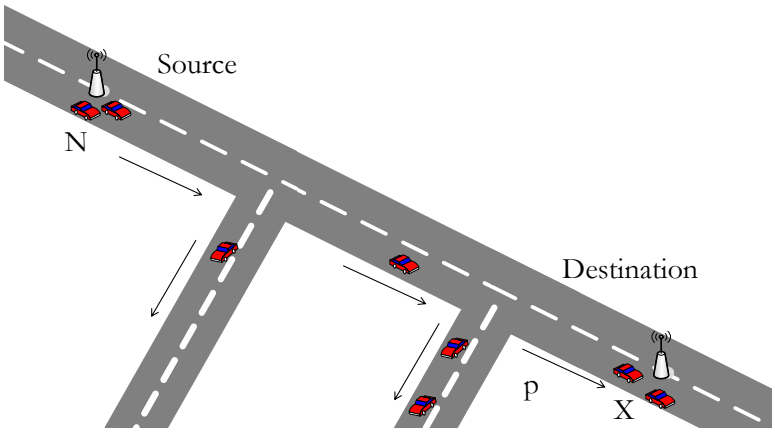


Figure 2.4: Number of relay vehicles depends on the probability of vehicles passing the destination. N is the total number of vehicles passing the source, X is the random variable representing the number of vehicles that have passed source will also pass the destination.

In this modified protocol, an RSU will relay a message to the minimum number of required vehicles; sometimes just one vehicle will be enough. However, this makes the protocol more prone to message dropping attacks (a malicious vehicle accepts a message for relaying but does not deliver it to the destination). Possible solutions include the use of end-to-end acknowledgement or an increasing of redundancy by using more than the minimum required vehicles for message relay.

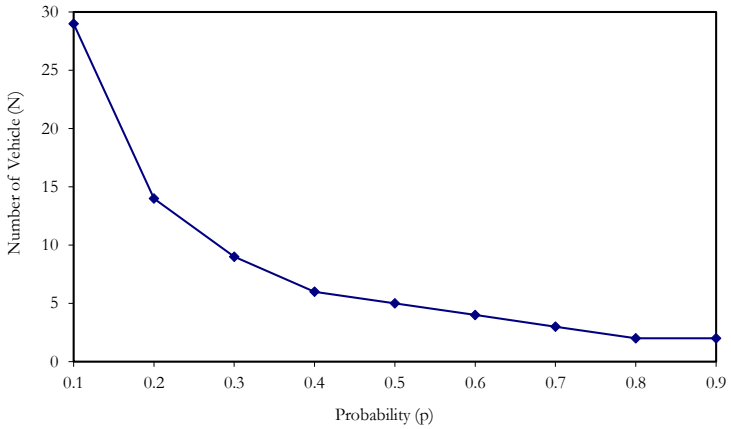


Figure 2.5: Number of relay vehicles (N) required to deliver a message to the destination (by at least one vehicle) with a 95% probability of confidence (Pc) for different probabilities (p) that a vehicle passing the source will also pass the destination. For example, if $p = 0.5$ and $Pc = 0.95$, we get $N=5$ which means that in order to have 95% confidence that a message sent by the source reaches the destination, we need to relay the message through at least 5 vehicles.

2.3 Simulations

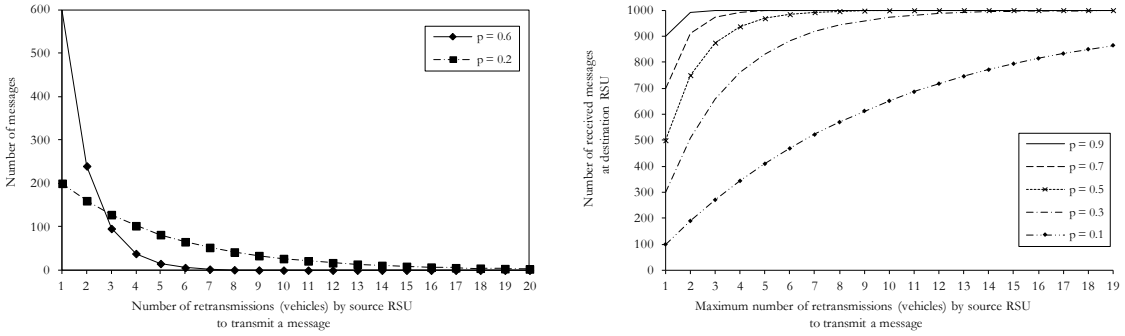
Simulations were carried out to check the effectiveness of our proposed system. The simulator does not incorporate the details of lower level protocol layers without the implementation of physical and MAC layers. All simulated vehicles and RSUs have the same transmission and reception ranges. A message transfer between a source and a destination is assumed to be successful if both entities are within the communication range of each other.

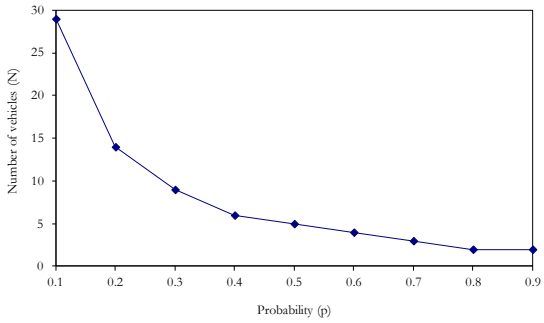
2.3.1 Simulation Scenario I

This set of simulations were carried out to find the minimum number of vehicles required to successfully transfer a message from a source RSU to a destination RSU with a given probability of confidence. A region of 25000m×6250m with road network as shown in figure 4 was simulated.

When a vehicle traveling towards the destination RSU passes the source RSU, the source RSU transmits the message to the vehicle. The message is then carried by the vehicle for possible delivery to the destination RSU. At each road junction, the vehicle decides to either maintain its direction of travel or divert according to a predefined probability. If the vehicle diverts and hence fails to deliver the message to the destination RSU, the source RSU retransmits the message. This procedure is repeated until the message is successfully received by the destination RSU. In each simulation run, the source RSU sends 1000 messages and the number of retransmissions for each message is recorded. The simulation is repeated 100,000 times and the average number of messages received successfully after a particular number of retransmissions is recorded.

Figure 2.6(a) shows the number of messages successfully received (Y-axis) using a particular number of retransmissions (X-axis) for $p = 0.2$ and $p = 0.6$ (p is the probability that a vehicle passing the source will also pass the destination). It shows that, for $p = 0.6$ case, more than 90% of messages can be successfully received by receiver within 4 retransmissions. From a different perspective, Figure 2.6(b) shows the number of received messages at the destination after less than or equal to each given number of retransmissions. Figure 2.6(b) can be used to find the minimum number of vehicles required to successfully transmit a message with a certain probability of confidence. Figure 2.6(c) shows the number of vehicles required for confidence Pc = 95% for each probability p. The simulation results shown in Figure 2.6(c) are identical to the analytical results presented in Figure 2.5.





(c)

Figure 2.6: (a) For the probability p=0.2 and p=0.6 (that a vehicle passing the source will also pass the destination), the number of messages successfully received at the destination RSU after a given number of retransmissions by the source RSU. (b) For different values of probability p, the number of received messages at the destination after less than or equal to each given number of retransmissions. (c) Number of Relay Vehicles (N) required to deliver the message to the destination with a 95% confidence probability (Pc) and different values of probability (p)

2.3.2 Simulation Scenario II

During the initial stages of VANET deployment, V2V communication will not be very effective. In addition, due to limited road infrastructure, the V2I communication will also be very limited. This will be a major setback to all VANET applications, such as transfer of a safety message from a point of incident to vehicles entering the area, or information about road blockage for possible diversion.

We have considered two cases and compared the number of vehicles and time required to transfer a message from a source of information (which can be a vehicle passing the scene of incident, or an RSU) to a destination (which can be an emergency response vehicle or an RSU). In the first case, we have a limited roadside infrastructure and messages are transferred between the source and the destination via vehicles only. In the second case, we have intermediate standalone RSUs between the source and the destination, which help in relaying the message. In this case the source is also a standalone RSU. Simulations will help us ascertain the effectiveness of our proposed system in relaying messages using vehicles with or without the intermediate standalone RSUs.

We simulate a region of 25000m×6250m with a road network as shown in figure 3. The number of smart vehicles on the simulation field (a total road length of 35000m), at any one time, is kept to 5. This small number of vehicles is used to check the effectiveness of our proposed system during the initial deployment stages of VANET. V2V communication is ignored due to this small number of smart vehicles. At each junction, a vehicle can divert from its current direction of travel with a probability of diversion Pd .

In both cases, the source RSU retransmits the message until it is received by the destination. In the second case, a vehicle carrying a message relays it to any intermediate RSU that it encounters. The number of retries (vehicles used to carry the information from the source) and the total time taken for the information to reach the destination is recorded for each message. A total of 1000 messages are transmitted in each simulation. The number of messages received at the destination after less than or equal to each given number of retries for $Pd = 0.5$ are shown in Figure 2.7(a). The results

indicate that the use of multiple (standalone intermediate) RSUs decreases the number of retries considerably.

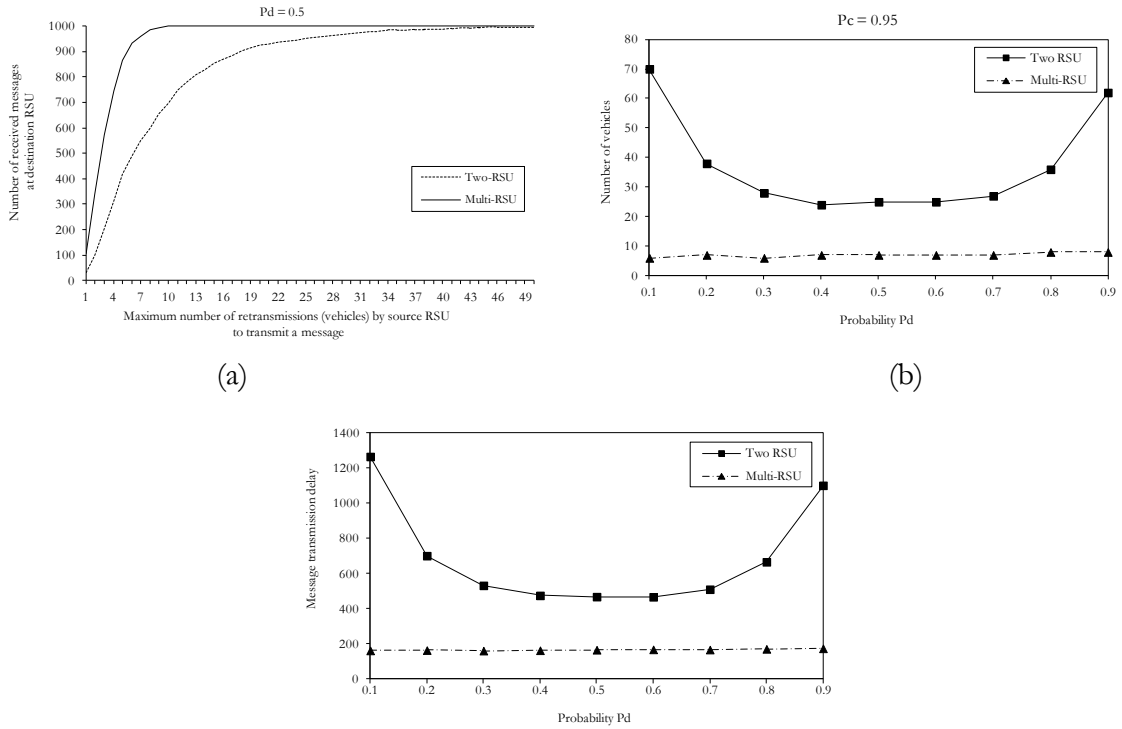


Figure 2.7 (a) For probability of diversion Pd = 0.5 (that a vehicle passing road junction will divert from its direction of travel), the number of received messages at the destination after less than or equal to each given number of retransmissions by the source RSU. (b) Number of Relay Vehicles used by the source RSU to deliver the message at the destination with a 95% probability of confidence (Pc) for different probabilities (Pd). (c) Message transmission delay for different probabilities (Pd).

Figure 2.7(b) shows the number of relay vehicles used to transmit the message to the destination with a 95% of confidence probability. The probability of diversion is varied from 0.1 to 0.9. The results show that for the first case (without intermediate RSUs) the number of vehicles reaches its minimum value when *Pd=0.5*. This is due to the road layout: at the first road junction a small value of *Pd* is helpful, but at the second road junction a large value of *Pd* is more advantageous. The number of vehicles required for the scenario with multiple RSUs almost remains constant. This

happens because the vehicles traveling on other roads also help in the successful delivery of messages. The same pattern of results is obtained in the transmission delay of messages as shown in Figure 2.7(c). The results indicate a high performance gain when multiple (standalone intermediate) RSUs are used, and the transmission delay will be much more stable than the case when only two RSUs are used. This is true for both the message transmission delays and the number of relay vehicles required for successful message transmission.

2.4 Discussions

The proposed system effectively meets the challenges highlighted in section 3. Details are given below:

- The proposed system provides an immediate solution to the problem existing during VANET initial deployment stage before a critical mass is achieved.
- The proposed system maintains VANET function in scenarios where V2V communication is not possible due to road layout or traffic conditions.
- The proposed system is progressive. RSUs of varying degrees of functionality can be integrated and later upgraded without the need to an overhaul of existing systems.
- The proposed system is an economical solution.
- The proposed system exhibits good scalability. More areas can easily be included in an existing VANET network by simply adding more RSUs. In addition, initially isolated regions can be later interconnected by RSU to RSU links.
- The minimum number of RSUs required for the proposed system to work is very small as compared to conventional solutions.

There are some limitations in the proposed design. First, because communication relies on RSUs to relay, it may be slow for vehicles to receive time-critical messages compared with V2V (or V2I with I2I) communication. However, in the VANET initial transition period, V2V and also I2I communication might not be possible due to the low density of smart vehicles on the roads and a lack of fully networked roadside units. Second, the VANET communication relies on the RSU infrastructure. It is possible that in some rural areas there are no RSU devices installed. Third, RSU to RSU indirect communication relies on passing-by vehicles. Thus the communication may be slow and can be interrupted frequently when there are few smart vehicles around.

2.5 Conclusion

There are numerous proposed applications of VANET but most of them are not practical until a critical mass of fully networked roadside units and smart vehicles is achieved. It will be very difficult to achieve this critical mass in the initial years of VANET deployment. This difficulty will further slow down the market penetration. In this paper, we have presented an economical and practicable solution to address this issue, which incorporates and relies on a very limited numbers of roadside units with very basic functionalities. Our solution is economical, scalable and upgradeable. We show that the solution is practical with the help from a small number of smart vehicles. The future work includes use of real traffic data for simulations and experiments

2.6 References

[1] A. Agarwal, D. Starobinski, and T. D. C. Little, "Exploiting downstream mobility to achieve fast upstream message propagation in vehicular ad hoc networks", INFOCOM/MOVE, May 2007.

[2] H. Wu, R.M. Fujimoto, R. Guensler, M. Hunter, "MDDV: Mobility centric Data Dissemination Algorithm for Vehicular Networks", VANET'04, Oct 2004.

[3] Q. Xu, T. Mak, R. Sengupta, "Vehicle-to-Vehicle Safety Messaging in DSRC", VANET'04, Oct 2004.

[4] G. Korkmaz, E. Ekici, F. Ozguner, U. Ozguner, "Urban Multi-Hop Broadcast Protocol for Inter-Vehicle Communication Systems", VANET'04, Oct 2004.

[5] A. Vahdat, D. Becker, "Epidemic Routing for Partially-connected Ad-Hoc Networks", Technical Report CS-2000-06, Duke University, Jul 2000.

[6] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in Sparse Vehicular Ad Hoc Wireless Networks," IEEE Journal on Selected Areas in Communications, vol. 25 issue 8, pp. 1538-1556, October 2007.

[7] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multi-copy case," IEEE Transactions on Networking, Vol 16, Issue 1, pp. 77 – 90, Feb. 2008.

[8] M. Mabiala, A. Busson, and V. Vèque, "Inside VANET: hybrid network dimensioning and routing protocol comparison", VTC'07-Spring,, Apr 2007.

[9] S. Jain, K. Fall and R. Patra, "Routing in a Delay Tolerant Network", SIGCOMM'04, Aug. 2004.

[10] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets", SIGCOMM'03, Aug 2003.

[11] T. Little and A. Agarwal, "An Information Propagation Scheme for VANETs," ITSS'05, Sep 2005.

[12] UMass DiselNet. http://prisms.cs.umass.edu/dome/umassdieselnet.

[13] Hull, B., Bychkovsky, V., Zhang, Y., Chen, K., Goraczko, M., Miu, A., Shih, E., Balakrishnan, H., and Madden, "CarTel: a distributed mobile sensor computing system", SenSys '06, 2006.

[14] D. Goodman, J. Borras, N. Mandayam, and R. Yates, "INFOSTATIONS: A New System Model for Data and Messaging Services", VTC'97, May 1997.

[15] T. Small and Z. J. Hass, "The Shared Wireless Infostation Model - A New Ad Hoc Networking Paradigm", MobiHoc'03, Jun 2003.

[16] M Bechler, WJ Franz, and L Wolf, "Mobile internet access in FleetNet", 13th Fachtagung Kommunikation in verteilten Systemen, Apr 2003.

[17] R. A. Wyatt-Millington, R Sheriff, Y. F. Hu, P. Conforto, and G. Losquadro, "The SUITED project: a multi-segment system for broadband access to Internet services", IEE Broadband Satellite Conf., 2000.

[18] J. Santa, R. T. Moreo, and A. F. G. Skarmeta, "A novel vehicle communication paradigm based on cellular networks for improving the safety in roads", Int. J. Intelligent Information and Database Systems, Vol. 2, No. 2, pp. 240-257, 2008.

[19] iCartel: MIT CarTel. http://icartel.net/icartel-docs/

[20] J. Gutiérrez. "Selected readings on telecommunications and networking", Idea Group Inc (IGI), 2008.

[21] Data Calculator: http://www.att.com/standalone/data-calculator

[22] J. Luo and J.-P. Hubaux, "A survey of research in inter-vehicle communications", in Securing Current and Future Automotive IT Applications, pp 111-122, Springer-Verlag, 2005.

[23] K Levacher, F McGee, F Murphy,"A comparison between 3G and 802.11 wireless technologies for Inter-Vehicular Communications purposes", http://killian.levacher.googlepages.com/Acomparisonbetween3Gand802.11wireles.pdf

[24] Y. F. Ko, M. L. Sim, and M. Nekovee, "Wi-Fi based broadband wireless access for users on the road", BT Technology Journal, vol. 24, pp. 122- 129, April 2006

[25] A. Qureshi and J. Guttag, "Horde: separating network striping policy from mechanism", MobiSys'05, Jun 2005.

[26] R. Chakravorty, A. Clark and I. Pratt, "GPRSWeb: optimizing the web for GPRS Links", MobiSys'03, May 2003.

[27] M. C. Chan and R. Ramjee, "TCP/IP performance over 3G wireless links with rate and delay variation", Mobicom'02, Sep 2002.

[28] J. Zhao and G. Cao, "VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks," INFOCOM'06, 2006.

[29] C. Lochert, B. Scheuermann, M. Caliskan and M. Mauve, "The Feasibility of Information Dissemination in Vehicular Ad-Hoc Networks", WONS'07, Jan 2007.

[30] B. Petit, M. Ammar and R. Fujimoto, "Protocols for Roadside-to-Roadside Data Relaying over Vehicular Networks", WCNC'06, April 2006.

[31] Y. Ding, C. Wang, and L. Xiao, "A Static-Node Assisted Adaptive Routing Protocol in Vehicular Networks," VANET'07, Sep 2007.

[32] M. C. Chuah, and F. Fu, "Performance Study of Robust Data Transfer Protocol for VANETs," LNCS - Springer, Vol 4325, pp. 377-39, 2006.

[33] A. Mansy, M. Ammar and E. Zengura, "Reliable roadside-to-roadside data transfer using vehicular traffic", MASS'07, Oct 2007

[34] M. Raya, P. Papadimitratos, and J.-P. Hubaux, "Securing Vehicular Communications," IEEE Wireless Communications Magazine, pp 8-15, 2006.

[35] P. Papadimitratos, L. Buttyan, J-P. Hubaux, F. Kargl, A. Kung, and M. Raya, "Architecture for Secure and Private Vehicular Communications," Intl. Conf. on ITS Telecom., Jun 2007.

[36] "IEEE P1609.2 trial-use Standard for Wireless Access in Vehicular Environments: Security Services for Applications and Management Messages," July 2006.

[37] F. Armknecht, A. Festag, D. Westhoff, and K. Zang, "Cross-layer privacy enhancement and non-repudiation in vehicular communication", WMAN'07, Mar 2007.

[38] C. I. Fan, R. H. Hsu, and C. H. Tseng, "Pairing-based message authentication scheme with privacy protection in vehicular ad hoc network", Intl. Conf. on Mobile Technology, Applications and Systems, Sep 2008.

[39] G. D. Crescenzo, T. Zhang , and S. Pietrowicz, "Anonymity notions for public-key infrastructures in mobile vehicular networks", MASS'07, Oct 2007.

[40] J. Choi and S. Jung, "A security framework with strong non-repudiation and privacy in VANETs", CCNC'09, Jan 2009.

[41] P. Papadimitratos, G. Mezzour, and J. P. Hubaux, "Certificate revocation list distribution in vehicular communication systems", VANET'08, Sep 2008.

[42] A. Stampoulis and Z. Chai. Survey of security in vehicular networks, project CPSC 534, http://zoo.cs.yale.edu/ ~ams257/projects/wireless-survey.pdf, 2007.

[43] N. Wisitpongphan, O. Tonguz, F. Bai, P. Mudalige, and V. Sadekar, "On the Routing Problem in Disconnected Vehicular Networks," INFOCOM'07, May 2007.

[44] IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Computer Society and the IEEE Microwave Theory and Techniques Society, (2004).

[45] S. Biswas, and R. Morris, "ExOR: Opportunistic multi-hop routing for wireless networks", SIGCOMM'05, Aug. 2005.

[46] J. Kim and S. Bohacek, "A Comparison of Opportunistic and Deterministic Forwarding in Mobile Multihop Wireless Networks", MobiOpp'07, 2007.

[47] V. Jacobson, "Congestion Avoidance and Control," In Proceedings of the ACM SIGCOM'88 Conference, pp. 314-329, August 1988.

[48] RFC 4346: "The Transport Layer Security (TLS) Protocol Version 1.1".

[49] Mobile Meter Reading, http://www.progress-energy.com/custservice/flares/meters/index.asp.

[50] J. Douceur, "The Sybil Attack", In First International Workshop on Peer-to-Peer Systems, Mar 2002.

[51] M. A. Lombardi, "National Institute of Standards and Technology (NIST) Special Publication 432", NIST Time and Frequency Services, Edition 2002 (revised April 2003)

CHAPTER 3 INTERNET ACCESS THROUGH SATELLITE RECEIVE-ONLY TERMINALS

Communication and especially the connectivity to the Internet is the basic requirement of most modern productive environments. We spend a considerable time traveling from one point to another via vehicles; this time can be more productive if we are connected to the Internet. A lot of research has been done to bring the Internet to vehicles. To this end, three main approaches have been adopted: Internet through roadside infrastructure alone or through roadside infrastructure using vehicle to vehicle (V2V) communication [1 - 7], Internet through cellular network [1, 8, 9] and Internet through satellite (symmetric/asymmetric) [10 - 14]. However, all three approaches have some challenges to deal with.

The Internet access through roadside infrastructure requires pervasive roadside units (RSUs) to achieve connectivity, since the typical radial range of an RSU is 250m so we need an RSU every 400 to 500m. Further, these RSUs must all be connected to the Internet. Also, the installation, connection and maintenance of these RSUs will be quite expensive, and it may not be possible to achieve the desired connectivity, especially during the initial days of VANET deployment and along highways or in rural areas. This approach is based on vehicular network architecture defined by IEEE standards, i.e., IEEE standards for Wireless Access in Vehicular Environments (WAVE) [15 - 19]. The Medium Access Control (MAC) and Physical Layer (PHY) of WAVE (IEEE P802.11p [19]) are based on *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY)* specifications

defined by IEEE standard 802.11 [20]. Both vehicles and RSUs use same access technologies specified by WAVE standards.

The Internet access through cellular networks, with the advent of 3G/4G technologies, provides data rates comparable to broadband; however, these technologies are not uniformly available throughout cellular coverage area, and hence, many users still get lower data rates. One of the design considerations for the cellular networks (to minimize the infrastructure cost) is to use larger cells where user density is low; this is especially the case along highways and in rural areas [21]. But in case of vehicular networks the density is likely to increase and may require up-gradation of cellular networks. In addition to these issues, when a vehicle crosses international boundary**,** the service providers and their carrier frequencies will often change, which makes user equipment more expensive and complex. Furthermore, this approach is also not compatible with the vehicular network architecture defined by IEEE standards for WAVE [15 - 19] since cellular access technologies are quite different from those specified by WAVE standards. This will make it expensive and difficult for both VANET spatial transition (between areas with varying degree of VANET coverage) and temporal transition (during later stages of VANET when previously not covered areas are also covered).

The Internet access through symmetric satellite requires vehicles to be equipped with satellite transceiver, which adds to the cost of user equipment. Satellite channel suffers heavily from losses; these losses introduce errors in the communication and require some error correction mechanisms. The losses can be atmospheric or due to shadowing. The atmospheric losses are generally compensated by link margin but shadow losses generally make communication impossible. These

shadow losses are much more pronounced in urban areas where the areas are congested with buildings and other manmade objects. Therefore Internet access through satellite is particularly not an economical solution for urban areas. Further, this approach is also not compatible with the vehicular network architecture defined by IEEE standards for WAVE [15 - 19], and will make both the spatial and temporal transitions difficult.

We present a solution that complements the existing ones and provides Internet connectivity during the initial deployment phase of the vehicular networks and also in areas with very scarce roadside infrastructure (such as along highways and in rural areas). The solution uses satellite receive-only terminals and very few (widely spaced) RSUs. It can support TCP connection even when the uplink is interrupted for long durations of time. We present a number of options with varying degrees of error handling capabilities and recommend their usage according to the environment. The use of RSUs makes our approach compatible to the vehicular network architecture defined by IEEE standards for WAVE [15 - 19]. Later on when additional RSUs are installed, the solution will improve its performance by making more use of RSUs and also by reduction of inter-RSU distance, and hence, enabling a smooth transition to fully functional VANET defined by IEEE WAVE standards [15 - 19]. Similarly, it will ease the handoff between areas of varying VANET coverage. The solution is cost effective, incremental and practical.

A lot of research to address challenges of Internet (especially TCP performance) over delay tolerant networks (DTN)/satellite networks has been carried out. In order to avoid the repetitions, we will, in this paper, not focus on lower level details of what particular Internet protocol to be used; rather,

we will identify the desired characteristics of the protocol and any already defined protocol (or combinations of these) can be used for the proposed solution.

The chapter is organized in 7 sections. Section 3.1 highlights the motivation and challenges related to the research work. Section 3.2 discusses related research work in the field. Section 3.3 explains important characteristics of satellite communication and mobile satellite communication model. Section 3.4 discusses the proposed system design and its various options along with the recommended usage. Section 3.5 describes the simulation scenarios and important results of the simulation. And in the end, section 3.6 presents conclusion.

3.1 Motivation and Challenges

If we define the utilization of a roadside unit as the length of time it successfully communicates with a particular vehicle (the utilization is directly proportional to the radial communication range of roadside unit and inversely proportional to the speed of the vehicle), then the utility of roadside units is much more in case of urban areas as compared to that in rural areas or along the highways. This is because of their complementary traffic characteristics, i.e., the average vehicle speeds are much lower in urban areas than that in rural areas or along highways, and also the ratio between move and stop is more inclined towards stop in urban areas (because of frequent intersections, turns and road signals) than that in rural areas or along highways.

This means the connectivity achieved with a given number of roadside units is much more in case of urban areas than that of rural areas or along highways. Therefore, to achieve same degree of

connectivity for a particular vehicle, we will need much more roadside units in rural areas (or along highways) than that in urban areas. Further, the installation and maintenance of roadside units in urban areas is much more economical than in rural areas. Also, the networking of roadside units to Internet is also easy in case of urban areas, since urban areas already have pervasive Internet connectivity. It is intuitive that the solutions completely based on roadside units are not practicable in rural areas (or along highways) especially during the initial deployment stages of vehicular network. Therefore, the extension of Internet to vehicular network in rural areas (or along highways) and especially during the initial deployment stages is not a trivial task.

The movement of vehicles is not restricted within one region; it is normal for users to travel through different urban/rural areas and along different highways on a single day. Any solution for rural area must be compatible with that of urban area: the urban VANET will, most likely, be based on RSUs as defined by WAVE standards. Therefore, the solution must support smooth spatial transition (handoff) between these regions. Further, during later stages as more and more area will be covered by RSUs, the solution should also support smooth temporal transition to mature VANET.

A number of digital video broadcasting (DVB) standards define interactive data services including Internet access via satellite, public switched telecommunication network, wireless etc [22 - 26]. In [23, 24], both the broadcast and interaction channels are via satellite which makes user equipment costly and incompatible with VANET standards. Whereas, in [26] both channels are via wireless (VHF/UHF bands) similar to VANET. DVB standards also support the combination of different DVB interactive systems [22]. Satellite as broadcast/downlink channel coupled with dial-up as interaction/return channel to provide Internet to home users especially in rural areas has been used

successfully for quite some time [28 - 32]. Our solution uses satellite as broadcast/downlink channel and terrestrial wireless (defined by VANET standards) as interaction/return channel.

The Internet and also some of the possible VANET applications exhibit asymmetric nature of traffic, in this downlink traffic is many orders of magnitude as compared to the uplink traffic [27, 28]. This asymmetry is likely to increase with time as more and more content is becoming multimedia in nature. We use asymmetric satellite communication (downlink only) to take advantage of this asymmetry. The use of satellite to provide connectivity in rural areas also seems logical since there will be less shadowing and hence low errors in rural areas (or along highways). The interaction/uplink is via roadside infrastructure using vehicle to infrastructure (V2I) communication or V2V in conjunction with V2I communication. However, the limited number of RSUs makes the traditional asymmetric satellite solutions impracticable. The challenge in VANET is the intermittent availability of terrestrial interaction/return channel with possible long disruption periods especially during the initial days of VANET deployment.

3.2 Related Work

The solutions presented so far for provision of Internet to vehicles can be broadly divided into three categories; first, the solutions relying on roadside infrastructure or vehicle to vehicle communication, second, the solutions relying in some way on cellular networks and third, the solutions making use of satellite links. We will refer some of the important research papers in these categories.

A number of researches such as FleetNet, Drive-thru Internet, etc extensively rely on road side infrastructure and/or vehicle to vehicle communication to provide Internet connectivity to vehicles [1 - 4, 6, 33]. The basic requirement for these solutions is availability of pervasive roadside infrastructure and/or a large number of smart vehicles. Both these assumptions are not realistic during the initial deployment stage, further use of vehicle to vehicle communication has many security issues, such as privacy, confidentiality, denial of service etc. Solutions based on existing WiFi networks face similar problems [7].

A number of researches have incorporated cellular networks in VANET [1, 8, 9, 34]. Cellular networks are mostly used as a backbone - a replacement to roadside infrastructure. Cellular networks, though pervasive, offer lower data rates as compared to Wi-Fi (roadside infrastructure). Although with the advent of 3G/4G technologies data rates close to broadband can be achieved, these technologies are not uniformly available throughout cellular coverage areas and many users are still dependent on other heterogeneous technologies (WAP, GPRS, EDGE, HSDPA, etc [35]). Further, cellular data plan subscriptions are expensive; an unconstrained plan with a 5GB/month limit costs approximately $700/year. 5GB/month means per day a user on average can send/receive 50 emails (20 with attachments), download a song and a game/app, view 40 web pages, posts 10 social media posts with photos, and watch a streaming video of 40 minutes [36]. Although unlimited data plans and those that cost tens of dollars are also available, but these plans have fine print conditions, such as 'usage patterns' (no file sharing, excessive usage, etc), 'can only be used on smart phones' (no tethering), 'can only access certain service' (email, predefined websites, etc), and 'must have a qualifying voice plan'. Some service providers are charging approximately $2/MB or 1¢/KB for web browsing. All major cellular service providers are now offering and encouraging users (such

as offering 'unlimited Wi-Fi usage with data plan') to access data through hotspots (which use Wi-Fi just like VANET roadside infrastructure instead of 3G/4G). This also highlights cost/benefit of Wi-Fi over 3G/4G. Cellular networks also have several other disadvantages such as expensive to built/maintain, billing/licensing issues among different service providers, higher roaming rates, large and variable latency, central switching/resource management, difficult to scale and occasional blackouts [5, 6, 37-42].

Use of satellite channel for provision of Internet to terrestrial (static) and mobile users has been an interesting topic of research. Most of the researches in this area are related to performance-studies or enhancements of Internet protocols over symmetric/asymmetric satellite channels with stationary nodes [28 – 32, 43 - 47]. Symmetric satellite communication does not take advantage of asymmetric nature of Internet traffic and are more expensive. Further, in this paper we are dealing with asymmetric satellite communication where the nodes are mobile nodes.

There are also quite a few researches dealing with the mobile nodes but most of these study Internet protocol performance/enhancements [10 - 14]. Further these consider symmetric satellite channel i.e., both uplink and downlink communication takes place via satellite. Symmetric communication requires expensive transceiver at the mobile nodes and it does not take advantage of the asymmetric nature of Internet communication. In this paper we are using satellite downlink communication only and the nodes are mobile nodes.

Our work comes closer to [48], where asymmetric satellite communication has been used for provision of Internet to the mobile nodes. In [48] the satellite is only used for downlink and uplink

is via cellular network. The system design requires the mobile node to be equipped with both the satellite and cellular interfaces. The design suffers from the disadvantages of using cellular network (described earlier). Also, the design does not incorporate any roadside infrastructure, which when available could provide much higher data rates at lower costs. Further this also implies that the design will not be very successful in urban areas since satellite communication is not very reliable in urban areas (connection/fade ratio can be 33.3/66.6 in higher density cities like New York [49]).

Our system design differs in a number of ways from the researches presented above. First, we use satellite communication for downlink only thus reducing complexity of user terminals and operating costs. We use roadside infrastructure for uplink communication and do not need any cellular transceivers or satellite transmitters at nodes. This eases compatibility with other vehicular network architectures. The design works with very small number of RSUs and is especially suited for initial deployment stages. We present a number of options with varying degrees of error handling capabilities and recommend their suitability for different environments.

3.3 Satellite Channel

3.3.1 Channel Characteristics

The satellite channel is characterized by long delays, high fading/attenuation to signal, high bandwidth and in-order packet delivery. As the signal travels from satellite to an earth station (or a mobile node as in this case) it undergoes a variety of impairments or losses. Some of these losses are constant, others can be calculated based on statistical data and some are dependent on weather

conditions [50]. A satellite system design takes care of all clear weather losses by including appropriate margins for these losses. It can be safely assumed that the existing satellite link takes care of all such losses, the only losses that need to considered are atmospheric attenuation (mainly due to rain, ice etc) and mobile channel losses (multipath fading and signal shadowing).

The multipath fading is caused because, the received signal in addition to direct signal also contains components which are reflected off different surrounding objects. These reflected or echo components mainly depend on the environment such as rural, urban, suburban etc. The fading can be short term usually caused by reflections over surrounding surfaces or long term caused by hills, buildings, trees etc [51]. The Shadowing occurs when the direct satellite signal is obstructed; main causes of shadowing can be buildings, trees, bridges etc. The shadowing also depends on the environment of the user. In case of geostationary satellites the shadowing and multipath fading is mostly determined by the user's mobility characteristics, the environment and satellite elevation angle [52]. So to summarize it can be said that the mobile channel losses are closely related to the environment of the user.

3.3.2 Channel Model

The effects of satellite communication on Internet and especially TCP have long been an area of Interest to researchers. For this different satellite channel models have been assumed/used. The simplest of these assumes satellite channel as an error-free/error-resilient channel and just studies the effects of delays, bandwidth and asymmetry [31, 47]. An extension to this model is to assume some fixed values of bit error rates (BER) and study the effects on protocols [46]. Another model is

based on additive white Gaussian noise (AWGN) satellite channel [43, 45]. All these models consider the receiver to be static, but in our case the node will be mobile and therefore the channel behavior will change with time. The most commonly used land mobile satellite channel (LMSC) model is a two-state Markov chain based channel model, which has been represented by a digital two-state Gilbert-Elliott model [52, 53]. In this paper we will use this two-state channel model.

It is a two-state ON/OFF model. In ON (1) state the communication is error free after applying existing satellite communication channel coding; the state mainly covers line of sight (LOS) region. In OFF (0) state communication errors are beyond the existing channel correction capability and reliable communication is not possible, the state mainly covers non line of sight (NLOS)/shadowed/deep-fade regions [12, 53, 54]. Figure 3.1 shows the two-state model [52]. Transition probabilities of this model depend on the environment (mean duration of ON/OFF state), vehicle speed and transmission (bit) rate [54]. The model excludes fading events with short durations, so the state transitions can be assumed to take place at cell boundaries, where a cell corresponds to a data segment.

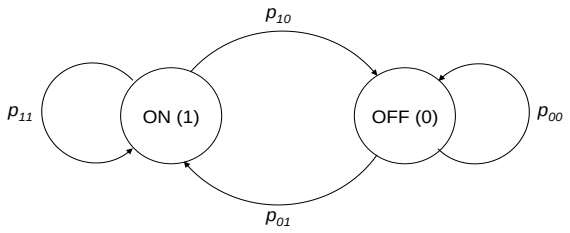


Figure 3.1: Two-state ON/OFF Land Mobile Satellite Channel (LMSC) Model

Denote p_{xy} as the transition probability of going from state x to state y, one-step state transition matrix is given by Equation 3.1. Denote π_x as the steady state probability of state x, the two steady state probabilities are given by Equation 3.2.

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \tag{3.1}$$

$$\pi_1 = \frac{p_{01}}{p_{01} + p_{10}},\ \pi_0 = \frac{p_{10}}{p_{01} + p_{10}} \tag{3.2}$$

Denote D_x as the mean sojourn duration of state x, for a constant transmission-rate/average-node-speed in a particular environment the two transition probabilities are given by Equation 3.3. Denote $p_x(>n)$ as the probability that a state x lasts for longer than n duration units, $p_x(>n)$ for two states are given by Equation 3.4.

$$p_{10} = \frac{1}{D_1},\ p_{01} = \frac{1}{D_0} \tag{3.3}$$

$$p_0(>n) = p_{00}^n,\ p_1(>n) = p_{11}^n \tag{3.4}$$

$$\textit{also}\quad p_{00} = 1 - p_{01},\ p_{11} = 1 - p_{10} \tag{3.5}$$

The average time in each state mainly depends on the environment in which the vehicle/node is moving, Table 3.1 shows the different environments, average vehicle speeds in these environments and average bad/good state times for these environments (the values of the Table 3.1 are from [54]), for more detailed explanation and data refer to [52]. (Table 3.2 summarizes various probabilities of the model).

Table 3.1 Average time in ON and OFF states for different environments

	Urban	**Suburban**	**Rural**	**Highway**
Vehicle Speed	Below 50Km/hr	Below 50Km/hr	Below 60-70 Km/hr	Below 120Km/hr
State ON (1) Duration – D_1	22s	8s	16s	18s
State OFF (0) Duration – D_0	15s	2s	4s	2s

Table 3.2 Probabilities associated to LMSC Model

Transition Probability from state x to state y	p_{xy} $p_{11}, p_{00}, p_{01}, p_{10}$
Steady state probability of state x	π_x π_1, π_0
Probability that a state x lasts for longer than n duration units	$p_x(>n)$ $p_1(>n)$, $p_0(>n)$

3.3.3 Satellite Communication

The satellite downlink communication makes use of existing error correction techniques on each transmitted segment. It is assumed that the existing error correction techniques applied are sufficient to provide error free communication in the absence of deep fading and shadowing [12, 53, 54]. To further reduce the effects of segment loss due to deep fading and shadowing, time diversity is applied [13, 55]. It can be achieved by inter-user or intra-user segment interleaving or both (Figure 3.2). This helps in spreading the error among different users or different sessions, and by employing error correction techniques at higher layer (discussed later) the chances of recovery are improved. The interleaving removes the impact of consecutive losses and therefore we can assume that

consecutive data segments of a session/user are independent of each other. This assumption simplifies our analysis of segment losses under different architectures that we propose next. In rest of the paper (especially in figures) the segments considered/shown adjacent to each other are consecutive segments of a session and are not necessarily transmitted consecutively unless described/shown otherwise.

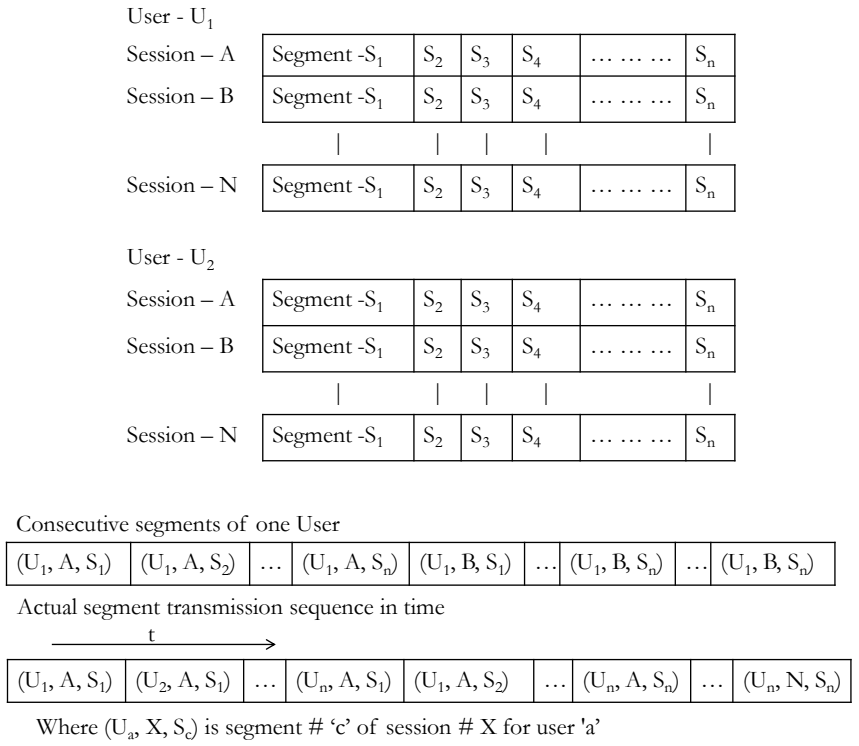


Figure 3.2: Inter and Intra user segment interleaving to achieve time diversity

Independence Analysis: We can provide mathematical formula to show the degree of dependency between two consecutive segments of one session's transmission after using the above segment interleaving method. The objective of interleaving method is to disperse a session's segments such that one satellite error will not cause consecutive losses for a single session. In term of the satellite

channel Markov on/off model, it means, we need to disperse two consecutive segments of a session such that their Markov states are independent.

Denote d as the number of Markov state transitions between the transmissions of two consecutive segments of a session. The Markov model d-step state transition matrix $P^{(d)}$ is:

$$P^{(d)} = P^d = \begin{bmatrix} p_{00}^d & p_{01}^d \\ p_{10}^d & p_{11}^d \end{bmatrix} \tag{3.6}$$

where p_{ij}^d is the d-step transition probability from state i to state j.

From the steady state analysis, we know that as the value of d increases, p_{00}^d and p_{10}^d will gradually converge to π_0, and p_{01}^d and p_{11}^d will converge to π_1. Therefore, we can define the degree of dependency after d-step state transition as:

$$P_{dependent} = \left(\frac{p_{00}^d - p_{10}^d}{\pi_0} \right)^2 + \left(\frac{p_{01}^d - p_{11}^d}{\pi_1} \right)^2 \tag{3.7}$$

The interleaving method requires that the degree of dependency between two consecutive segments of a session is smaller than a predefined parameter δ, i.e., $P_{dependent} < \delta$. To achieve this, we can adjust the interleaving parameter, d, to satisfy this requirement.

Because the *d*-step transition probabilities in Equation 3.7 do not have closed form solutions, to find the suitable value of *d*, we could try *d=1, d=2,* … to derive the corresponding values of $P_{dependent}$, until the requirement $P_{dependent}<\delta$ is satisfied.

3.4 Proposed Architecture

In this paper our focus is on provision/extension of Internet to vehicular networks in rural areas and along the highways especially during the initial deployment stages. The working environment is characterized by a very small number of RSUs that are widely interspaced. These RSUs may be co-located with isolated populated areas along the highways and are connected to the Internet. The environment does not exhibit high shadow losses.

3.4.1 Assumptions

Our system design is based on a few simple assumptions. First, vehicles are equipped with GPS, can record their location at precise time and can provide direction of travel information to the RSU. Second, vehicles can receive the satellite broadcast. And third, RSUs have the digital map of the area and are aware of the locations of adjacent RSUs.

3.4.2 Basic Idea

A vehicle connects to a nearby RSU and requests some Internet data. The request will include location, speed and direction of travel of the vehicle. This information will help an RSU to calculate

possible connection time left and possible next RSU. If sufficient connection time is left then the request may be serviced through the same RSU. When the vehicle exits the coverage of current RSU, further responses to the vehicle's earlier request will be sent to the next RSU in the direction of travel. If the two RSUs are located at a reasonable distance, which is likely in urban environment, the next RSU will continue to deliver the content to the vehicle when the vehicle comes within its coverage area (RSU-based region in Figure 3.3). If the next RSU is not within a reasonable distance (especially in rural environment where the RSUs will be widely spaced) then satellite downlink channel will be used for delivery of content (satellite downlink-only region in Figure 3.3). We will mostly address the satellite downlink option in this paper.

While a vehicle travels between two RSUs in the satellite downlink-only region, it cannot send acknowledgements. In order to keep TCP connection alive and avoid unnecessary retransmissions, adaptive TCP timeout and delayed ACK will be used [56, 57]. TCP timeouts will be calculated/predicted depending on the location of the next RSU and will be used accordingly. The downlink has large delays so in order to avoid unnecessary retransmissions selective acknowledgement will be used. Modified TCP is only employed between the proxy and mobile host so no modifications are required in protocols running on existing Internet. The use of UDP is much simpler than TCP and will not require any modifications. The protocol stacks are shown in Figure 3.4.

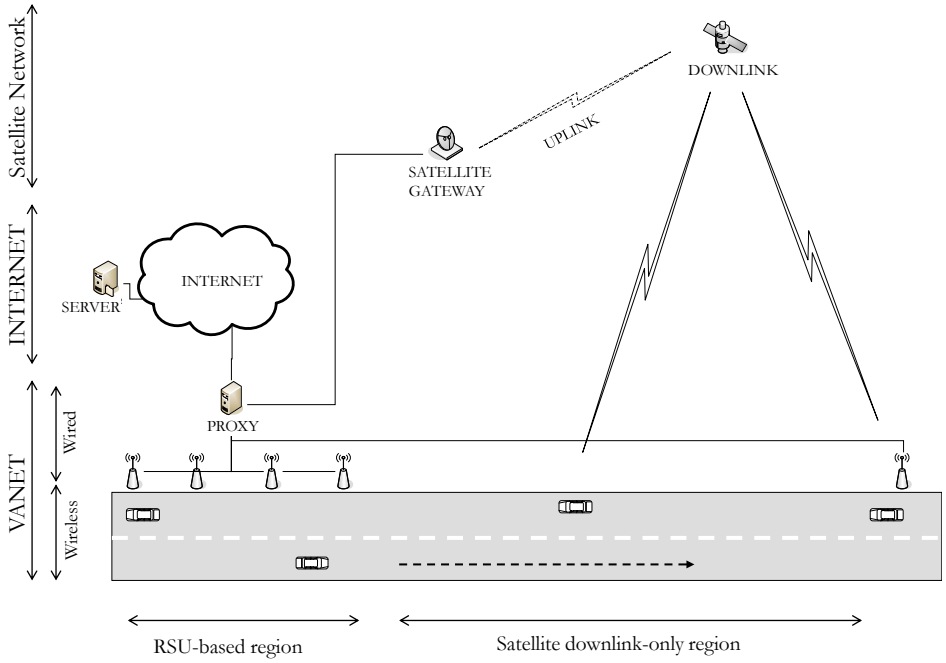
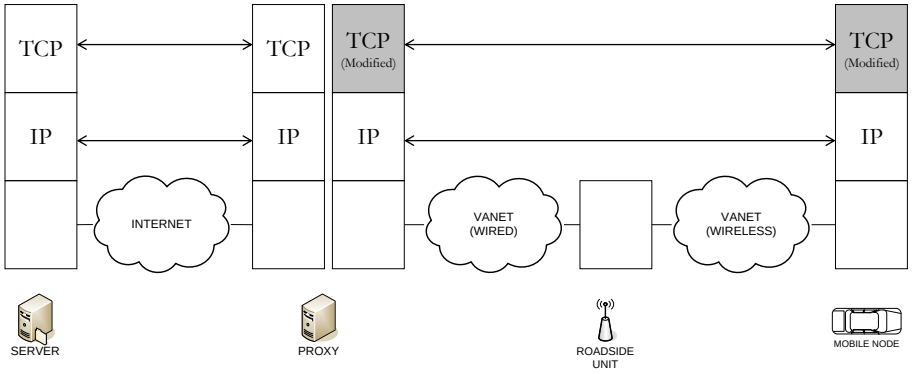
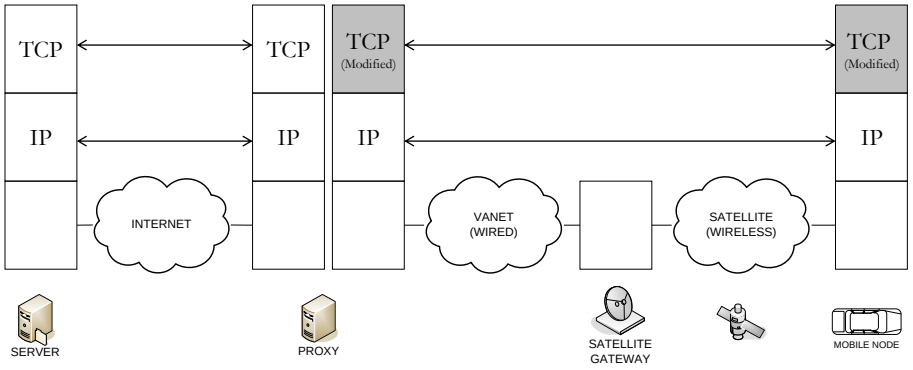


Figure 3.3: Proposed design in context of higher level vehicular network architecture



(a)



(b)

Figure 3.4: Protocol stacks for different regions. (a) Content transfer to mobile node via Road side units. (b) Content transfer to mobile node via satellite.

The flow of traffic between different entities is outlined below (refer to Figure 3.3):

- The mobile node authenticates with the proxy through an RSU and is issued with an IP address; this IP address will uniquely identify the mobile node as long as it remains within the boundary of the proxy.
- The mobile node sends a request to the RSU.
- The RSU forwards the request to the proxy.
- The proxy establishes a connection to Server on behalf of the mobile node and gets/caches all the content (based on initial request of the mobile node).
- The proxy establishes a connection with the mobile node on behalf of Server and sends the content via an RSU.
- When the mobile node moves out of the range of an RSU and the direct connection with RSU times out (Figure 3.3, satellite downlink-only region); the RSU informs the proxy about expected time of next ACK from the mobile node. The time period depends on speed of the vehicle and location of the next RSU along the travel direction of the mobile node.

- The proxy starts sending further content via satellite. It keeps on sending without waiting for ACK from the node till the time period expires.
- The mobile node keeps on receiving data via satellite till it reaches the next RSU (Fig. 3, end of satellite downlink-only region).
- The mobile node sends ACK for the received data or NACK for segments lost due to errors.
- The proxy updates the mobile node's new position and acts according to ACK/NACK.
- The process is repeated till all the requested content is delivered to the mobile node.

3.4.3 Options

Four different options have been defined for transmission of data from satellite to vehicles: Baseline, Repeated Transmission, Forward Error Correction, and Error Location Prediction and Avoidance. These options have different levels of error handling capabilities with corresponding overheads and delays. A comparison between these options is given in Table 3.3.

3.4.3.1 Baseline

A vehicle sends a request to its nearby RSU (Figure 3.5, R_1), which in turn forwards the request to the proxy server. The proxy server gets the response/data from the server and forwards it to the satellite gateway. The proxy server splits the end-to-end connection between the vehicle and server [58]. It maintains two separate connections, one with the server on behalf of the vehicle and the other with the vehicle. The session with vehicle will be asymmetric, that is, the down link will be

through satellite and return will be through RSUs. No modifications are required on the server side nor on the satellite downlink.

The connection between the proxy and the vehicle will employ TCP enhancements/modifications such as Adaptive timeout, delayed ACK, and selective ACK/NAK [56, 57]. The adaptive timeout caters for the time during which the vehicle cannot send ACKs, that is while traveling between the RSUs (Figure 3.5, between R_1 and R_2).

When a request is received by the proxy, it forwards the request to the original server in a separate connection. It also calculates the timeouts and delays (for delayed ACK) expected based on the distance between adjacent RSUs (Figure 3.5, between R_1 and R_2), the vehicle speed and its direction of travel. The proxy server receives all the data, which maybe a large file, from the server and also keeps the connection alive for further requests from the vehicle (this will be necessary if the transaction has to be completed after receiving some response from the vehicle during its connection with the next RSU). The proxy server then forwards the data to the vehicle through the satellite gateway and waits for the ACK/NAK. Because of high bandwidth-delay product (the delay of the satellite communication and also between sending the data and receiving ACK due to separation of RSUs), it may be necessary that all data segments are sent before waiting for an ACK/NAK from the vehicle.

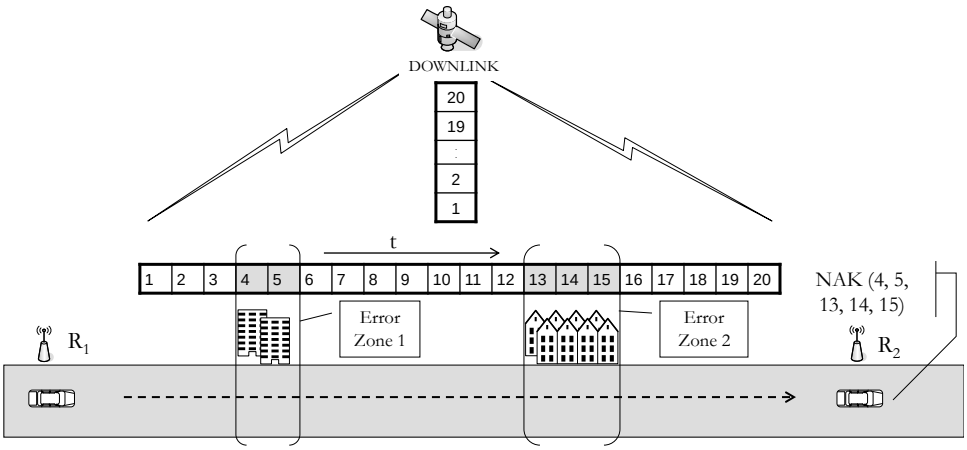


Figure 3.5: Baseline architecture, delivery of content takes place through satellite while the mobile node is traveling between R_1 and R_2. The mobile node sends NAK for the lost segments when reaching R_2, where R_2 takes charge and resends these lost segments to the mobile node.

If some of the received frames have been lost (Figure 3.5, segments 4, 5, 13, 14 and 15) then the vehicle sends selective ACK/NAK on its next contact with roadside infrastructure (Figure 3.5, R_2). These ACK/NAK segments are forwarded to the proxy server, which retransmits the lost segments through satellite/RSU.

Analysis: If the transmission consists of N chunks of data segments and each chunk is transmitted independently, the probability p that a chunk is transmitted during Good state is:

$$p = \frac{\pi_1}{\pi_1 + \pi_0} = \frac{p_{01}}{p_{01} + p_{10}} \tag{3.8}$$

The probability that all N chunks of data segments are successfully transmitted, before reaching the next RSU, is given by:

$$P_{success} = p^N \tag{3.9}$$

3.4.3.2 Repeated Transmissions

A method to address segment losses is by repeating the complete transmission in cyclic manner for a fixed number of times. This option adds maximum data redundancy. Although this is not an efficient utilization of the bandwidth available and we will have low information per bit transmitted, this approach can mitigate the effects of channel impairments. Especially during the initial VANET deployment stages when not many of smart vehicles will be on roads, a given satellite channel will be shared by a limited number of vehicles and each vehicle will have sufficient share of satellite bandwidth, which can be used for repeated transmissions. Also, during the initial stages there will be fewer number of RSUs which means larger distances between RSUs and more time to service a given request. This available time can be utilized for the redundancy.

This scheme suffers from long delays because in worst case a vehicle might have to wait for a complete cycle of retransmission to recover the lost data segment. It has high delay but is a suitable scheme when we are experiencing high error rate that cannot be corrected by other schemes such as Forward Error Correction (FEC).

Figure 3.6 shows a vehicle driving between two widely spaced RSUs. The data is being sent through satellite which comprises of segments numbered 1 to 8. The vehicle fails to receive segments 4 and 5 during the first transmission cycle since it was passing through error zone 1 during their transmissions. The vehicle recovers the lost segments from the second transmission cycle and is successful in receiving all eight segments before reaching the next RSU, where it acknowledges the receipt of all the segments.

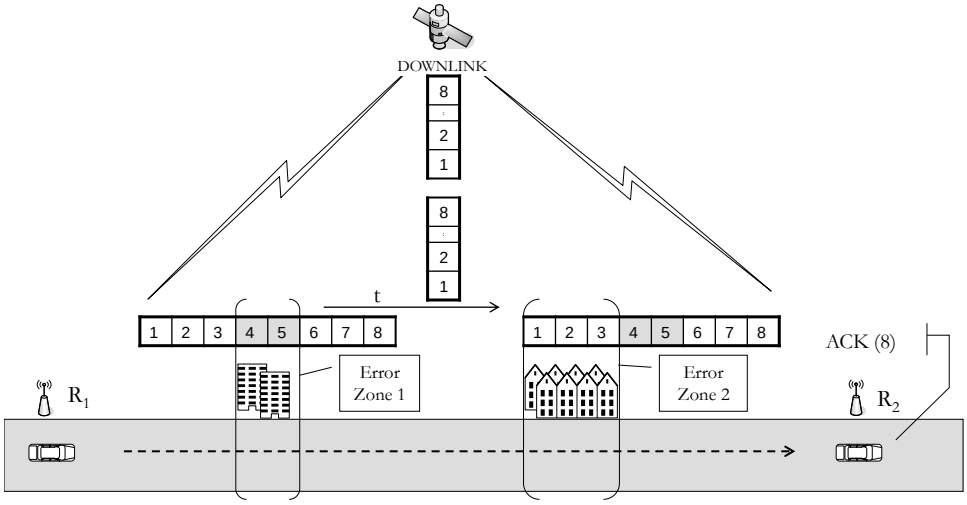


Figure 3.6: Repeated transmission, whole data set is repeated several times. Data segment(s) lost can be recovered from later repeated transmissions.

Analysis of the number of transmissions: We present the mathematical formula to derive the number of transmissions required in order to achieve a desired transmission success rate. This analysis is possible because the segments of one session are transmitted independent with each other as introduced earlier.

Assume that a session transmission requires N chunks of data segments. Because each chunk is transmitted independently, it is transmitted during the satellite Good signal status with the probability p as defined in Equation 3.8).

Denote the system transmission success probability when it implements n times transmission as $P_{success(n)}$. The session transmission is treated as successful if the recipient receives all N chunks of data without error before reaching the next RSU. Thus we can derive Equation 3.10. If the desired

transmission success rate is required to be *Pr*, we need to choose the number of transmissions *n* such that $P_{success(n)} > P_r$. $P_{success(n)}$ for $n = 2, 3$ and different values of N are shown in Figure 3.9(a).

$$P_{success(n)} = [1-(1-p)^n]^N \quad (3.10)$$

3.4.3.3 Forward Error Correction (FEC)

When deep fades happen, complete segments may be lost so link layer error correction mechanism implemented by satellite downlink cannot recover all lost data. For such a situaton, FEC and interleaving at higher layers such as transport layer may be used to address the issue of long burst errors. Two possible schemes can be adopted, FEC at session level and FEC at segment/packet level [51]. Both have their advantages and disadvantages.

At segment level, a fixed number of data segments are grouped together and parity segments are added based on the selected FEC algorithm. The success depends on the number of segments lost. In case of erasure codes the lost data can be recovered as long as the number of lost segments is no more than the number of parity segments added [59].

At the session level, the session is first divided into fixed sized blocks and then parity blocks are added to it. Since all sessions are not of equal size, FEC codes that operate on a fixed number of blocks cannot be employed here. Different sessions can be further interleaved to further spread each individual session in time and avoid the damages of long burst of errors.

The segments for different users, vehicles in this case, can be further interleaved or multiplexed. This provides another layer of spreading in time and the number of segments lost of a particular node or of a particular session will be further reduced by a factor defined by the number of nodes. This makes error correction mechanism more robust to deep fades and shadow. It may be possible to limit the number of lost segments per session within the tolerance of FEC employed. If the number of segments lost is more than the tolerance then the lost segments may be requested again at the next RSU through NAK mechanisms explained earlier. It is important to note that retransmission of all lost segments may not be necessary in this case since we only need to bring the number of lost segments within the tolerance range of FEC employed.

Analysis: If the session/file consists of N data segments and α rate erasure codes are used, aN segments will be added to the transmission and can therefore handle up to aN segment losses. The success probability, before reaching the next RSU, can be given as Equation 3.11 which mean we can derive Equation 3.12. $P_{success}$ for α = 0.10, 0.20 and different values of N are shown in Figure 3.9(b).

$$P_{success} = prob\ (\leq aN\ segments\ are\ bad) \tag{3.11}$$

$$P_{success} = \sum_{i=0}^{\alpha N} \binom{(1+\alpha)N}{i} (1-p)^i\, p^{(1+\alpha)N-i} \tag{3.12}$$

3.4.3.4 Error Location Prediction and Avoidance

The satellite mobile channel suffers mostly from shadowing and fading. These errors are strongly correlated to the environment. Vehicles traveling along a particular highway are expected to experience channel impairments at approximately the same locations (referred as "error zones"). If

the location of these error zones can be registered and the segments that were sent to a vehicle while it was passing through these error zones can be determined, then these segments may be retransmitted to the vehicle without waiting for an ACK/NAK from the vehicle. This will improve the performance since the vehicle does not need to wait for the next RSU, which may be quite far off, to recover from the error.

The location of error zones can be determined if a vehicle also includes location information with NAK, which is the location where segment loss was experienced. This location information is used to predict the possible location of segment losses for future vehicles traveling along the same path. For example in Figure 3.5 the vehicle experiences loss of segments 4 and 5 in error zone 1 and can report {NAK (4,5), Error Zone 1} to RSU-R_2. (Also loss of segments 13, 14 and 15 in error zone 2 can be reported as {NAK (13,14,15), Error Zone 2}).

The error locations reported by different vehicles may have some variations; these variations may be due to a slight shift in error location, imperfections in recording error location, or imperfection in knowing packet loss locations. The effects of this variation among different error zone locations can be minimized by using a smoothed error location (similar to TCP RTT model [60]). If α is the smoothing factor (that determines the weightage of old value) and ε is the latest reported error location, the smoothed error location E can be defined by Equation 3.10. The error zone span Z can be taken as $Z = C\,V$ centered at E, where V is the bad-state/error-zone variance and C is a constant.

$$E = aE + (1-a)\varepsilon \tag{3.13}$$

The time period during which the vehicle was traveling through a registered error zone can be estimated from the vehicle's speed and its initial time-location information. The fact that the speed of vehicles has generally less variations in highway environment also helps in minimizing the estimation error. However, an error margin can be added on both sides of probable error location to cater for variations in driving speeds. The RSU with which the vehicle was last authenticated/associated (Figure 3.7, R_1) listens for the segments which were sent during the error zone period and sends NAK to the proxy for retransmission of these segments.

Take Figure 3.7 as an example. A vehicle sends a request to RSU R_1, which forwards the request to the proxy server and data is sent to the vehicle via satellite. The RSU R_1 calculates approximate times ($\{t_0\text{-}t_1\}$ and $\{t_2\text{-}t_3\}$) when the vehicle will be passing through error zones. It then monitors the satellite transmissions destined to this vehicle. It records the segment numbers sent by satellite when the vehicle passes through these error zones (segments {4, 5}, {11, 12, 13}) and sends NAK to the proxy server for these segments (segments NAK {4, 5}, NAK {11, 12, 13}). When receiving these NAKs, The proxy server retransmits the lost segments through satellite.

When the vehicle reaches the next RSU, it sends NAK if it still could not receive all the segments. These NAK locations are then used by previous RSU to modify its error zone information. It is important to note that there is no way for the initial RSU to know if an earlier reported error zone has disappeared or not. There may be a situation (e.g. heavy rain) when most of the locations are marked as error zones. To address this possible situation it is necessary from time to time to reset the error zones to zero.

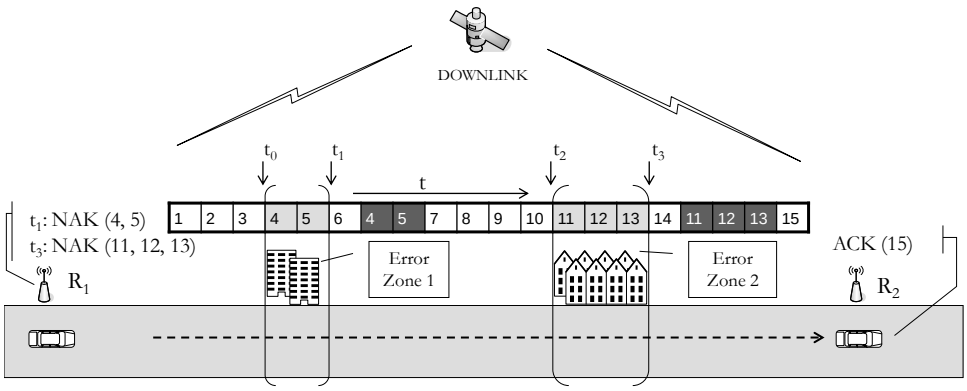


Figure 3.7: Error location prediction and avoidance. The system predicts the segments which may have been lost on the bases of previous data and proactively retransmits these segments.

Analysis: If a session consists of N data segments and the segments transmitted during Bad states that are successfully predicted (assume p_d is the probability for successful error zone prediction) are retransmitted during a future Good state. Then the probability that all N segments are successfully transmitted, before reaching the next RSU, is given by Equation 3.14. $P_{success}$ for p_d = 0.07, 0.08, 0.09 and different values of N are given in Figure 3.9(c).

$$P_{success} = [p / (1 - p_d(1-p)]^N \tag{3.14}$$

3.4.4 Enhancements Using V2V Communication

V2V communication can be used in a variety of ways to further enhance the efficiency of above mentioned techniques.

3.4.4.1 Local Error Recovery

If same satellite channel is used by a number of vehicles in a region in time sharing basis then it is possible for vehicles to cache the data destined for other vehicles. The amount of data cached and how long it is cached is function of storage space available. A cyclic buffer may be used for this purpose, when it becomes full the oldest data is overwritten. Fast indexing can be done using hash tables. This way when a vehicle exits error zone, it can sent NAK to next coming vehicle which may be able to transmit the requested packet from its cache (Figure 3.8a, 3.8b, 3.8c).

3.4.4.2 NAK/ACK Relay

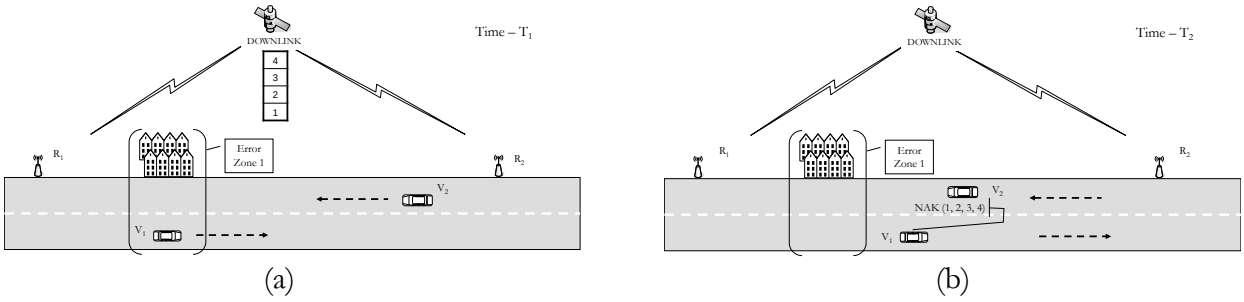
Another possible use is to relay NAK to previous RSU using vehicles traveling in opposite direction. This way the lost packet can be retransmitted via satellite and the vehicle does not have to wait till it reaches next RSU to report lost packets (Figure 3.8a, 3.8b, 3.8d). Further extension of this approach could be to use V2V communication as reverse channel to send all the selective ACKs and NAKs.

3.4.4.3 Retransmission Relay

The retransmission of lost packets may also be carried out by next RSU via vehicles traveling in opposite direction (Figure 3.8a, 3.8b, 3.8d, 3.8e). In this case the security of spoofed acknowledgements has to be ensured.

3.4.5 Comparison of Options

The options presented in preceding sections offer different levels of error tolerance at the cost of overheads and delays. One must balance the performance (error tolerance) vs. cost (overheads and delays) in selecting a particular solution; also, some options may be more suited to a particular environment than the other environments. The baseline architecture uses simple ACK/NAK for flow control and error correction. This scheme has no overhead but successful completion of communication may be delayed till the vehicle reaches the next RSU. This architecture is suitable where RSUs are not very widely dispersed. Error location prediction and avoidance uses proactive retransmission of predicted lost segments. This scheme has low overheads and low delays. This scheme is especially useful where satellite mobile channel losses are reasonably localized in certain areas. Forward error correction based architecture has low delays at the cost of medium overheads. It is suitable where errors are randomly distributed and Bad state durations are within the FEC tolerances. Repeated Transmission is the most robust scheme, it is especially useful where longer durations of Bad state is experienced, but at the cost of having high overhead and medium delays. Figure 3.9(d) compares the success probabilities of different architectures. Note that the error location prediction (p_d=0.9) performs almost the same as repeated transmission (n=2). A summary of the options with recommended usage is also presented in Table 3.3.



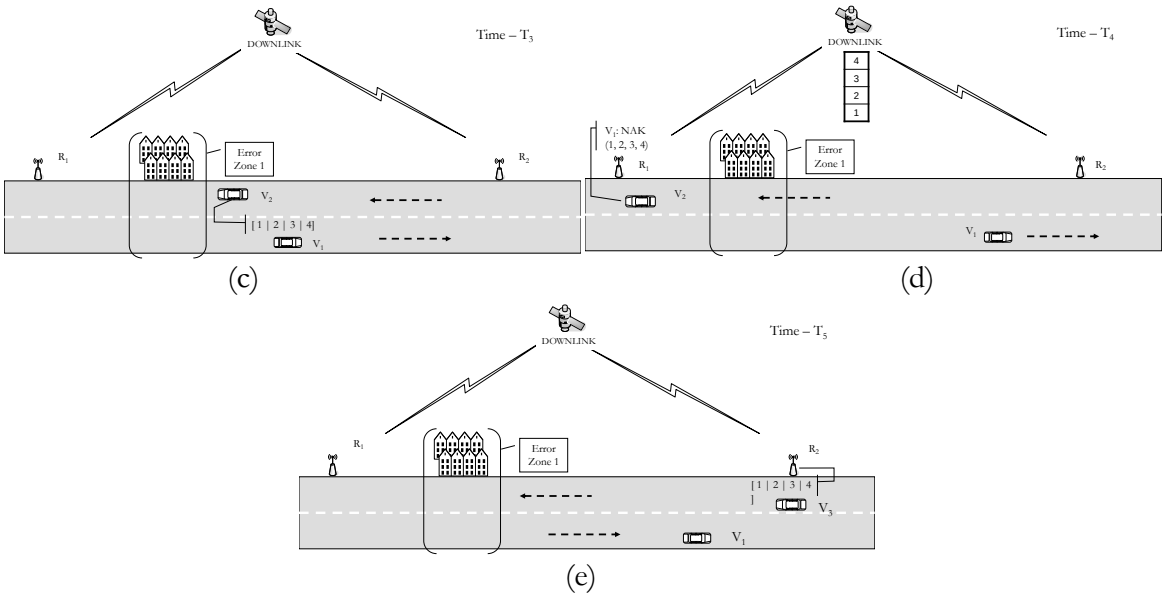
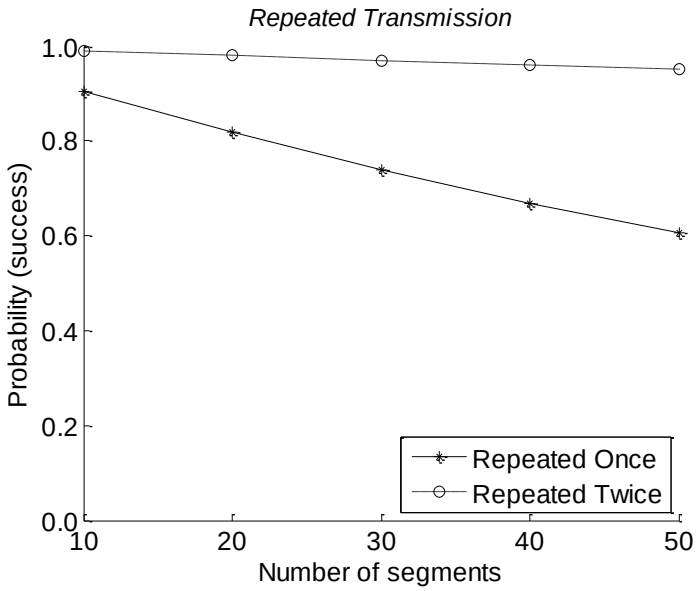


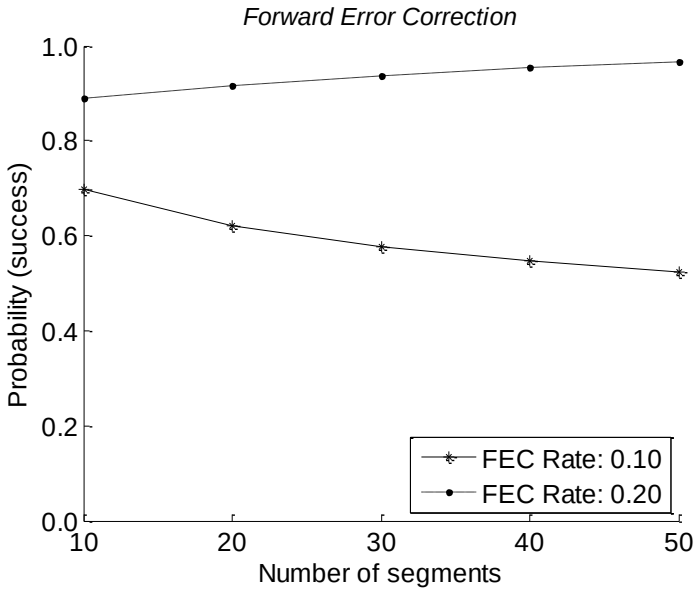
Figure 3.8: Enhancements using V2V communication. (a) packets {1,2,3,4} are lost by V_1 since traveling through error zone. (b) V_1 send NAK{1,2,3,4} to V_2. (c) V_2 sends the cached packets {1,2,3,4} to V_1. (d) NAK{1,2,3,4} of V_1 are relayed to R_1 by V_2, and the packets{1,2,3,4} are retransmitted via satellite. (e) Packets {1,2,3,4} are being relayed to V_1 by R_2 via V_3. Local Error Recovery: (a)→(b)→(c), NAK/ACK Relay: (a)→(b)→(d), Retransmission Relay: (a)→(b)→(d)→(e)

Table 3.3 Recommended usage of different options

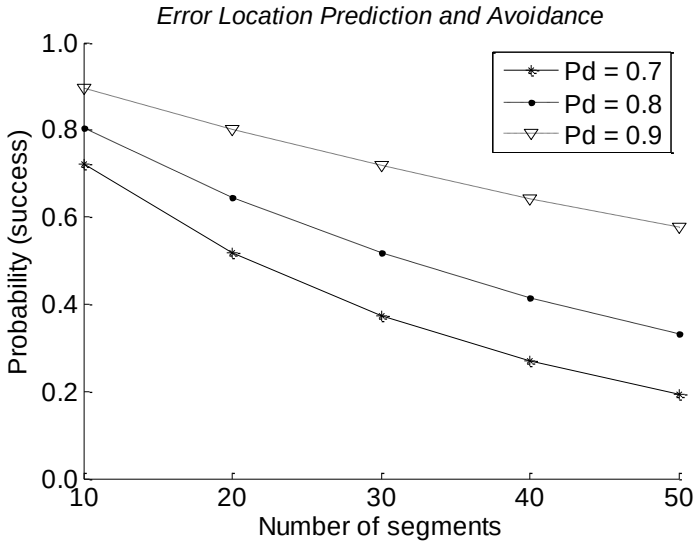
Option	Overhead	Delay	Recommended Usage
Baseline	None	High	Where RSUs are not very widely spaced
Repeated Transmission	High	Medium	Where Bad state duration is longer than FEC tolerance
Forward Error Correction	Medium	Low	Where Bad state duration is within FEC tolerance
Error Location Prediction and Avoidance	Low	Low	Where Bad state environments are relatively stable over a relatively long period of time


Repeated Transmission
Probability (success)
1.0
0.8
0.6
0.4
0.2
0.0
10
20
30
40
50
Number of segments
Repeated Once
Repeated Twice

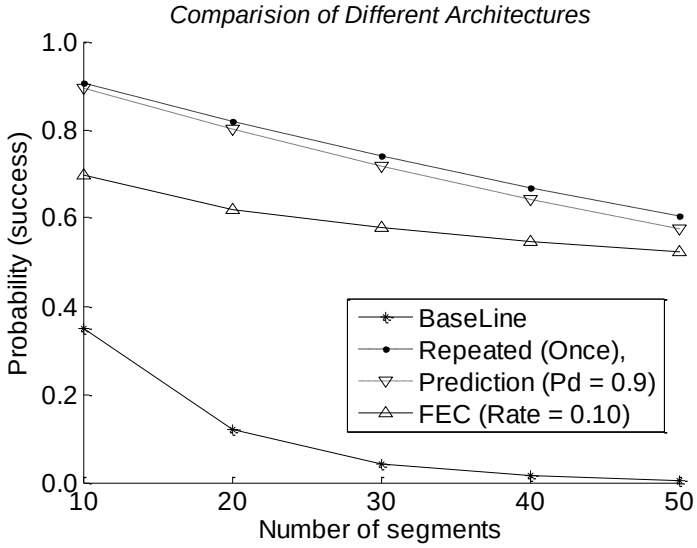
(a)


Forward Error Correction
Probability (success)
1.0
0.8
0.6
0.4
0.2
0.0
10
20
30
40
50
Number of segments
FEC Rate: 0.10
FEC Rate: 0.20

(b)



(c)



(d)

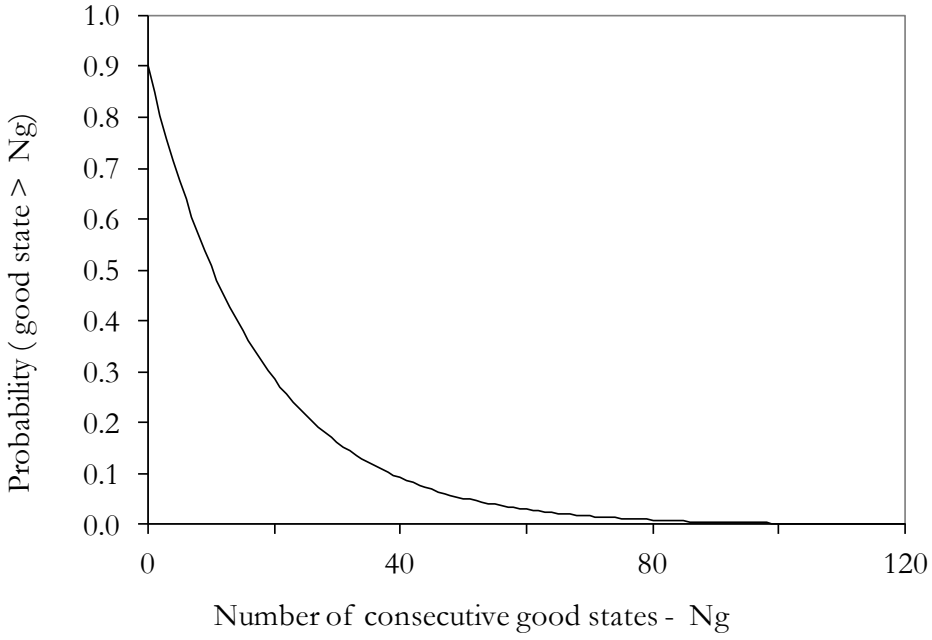
Figure 3.9: Analytical success probabilities (a) Repeated transmission (n = 2, 3) (b) Forward error correction (α = 0.1, 0.2) (c) Error location prediction and avoidance (p_d = 0.07, 0.08, 0.09) (d) Comparision between baseline, repeated transmission, forward error correction and error location prediction & avoidance architectures.

3.5 Evaluation

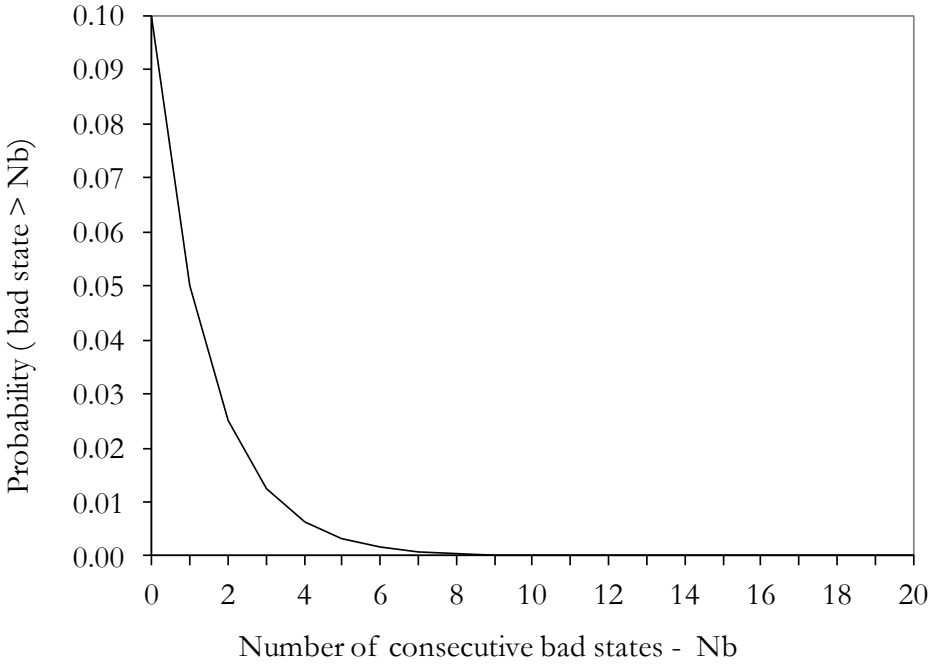
Simulations were carried out to ascertain the correctness of our analysis and to compare the performances of various presented architectural options. It is assumed that the node/vehicle has already requested the file/content via an RSU and is now travelling through the "satellite downlink-only region" where, the requested file/content is being sent via satellite downlink (refer to Figure 3.3). We mainly address the satellite downlink part in this paper, and hence, the simulations were also restricted to "Satellite downlink-only region". The satellite downlink has been modeled according to the land mobile satellite channel (LMSC) model presented in section III. The environment is Highway with parameters defined in Table I. The simulated LMSC characteristic probabilities are shown in Figure 3.10, which closely match the ones defined by Equation 3.2, 3.3 and 3.4. 100 sets of Markov chains were generated with equal distribution of initial/starting state. Each Markov chain had a length of $2x10^6$. For simulation first $5x10^5$ states were skipped to offset the effect of initial/starting state of a particular Markov chain.

The file size is defined in terms of segments and varies from 10 to 50. It is assumed that one segment is sent during one Markov state, each state being of 1ms. A segment sent during a Good state is received error free while a segment sent during a Bad state is lost. The time between two consecutive segments of a particular node/vehicle follows Poisson distribution with parameter λ. 100,000 simulation runs were carried out for each file; every 1000 simulation runs used one generated satellite channel Markov chain model. The simulation results compared with the corresponding analytical results are shown in Figure 3.11. This figure confirms the correctness of analysis presented earlier. The performance of Error location prediction and avoidance at Pd=0.9 is

comparable to Repeated transmission but with a much lower overhead. The transmission time and segment loss probabilities for different architectures are shown in Figure 3.12; (a) gives the probability that the segments lost (measured as a fraction of total file size) is less than or equal to a given value and (b) gives the probability that the file transfer time (normalized with the mean transfer time of baseline architecture) is less than or equal to a given value. The results clearly show the superiority of Error location prediction option over Repeated transmission and also confirm the characteristic descriptions of different architectural options presented in Table 3.3.

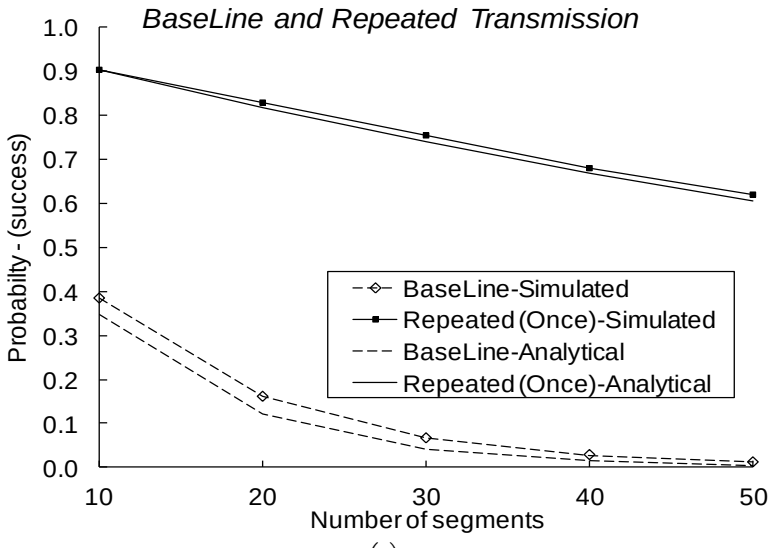


(a)

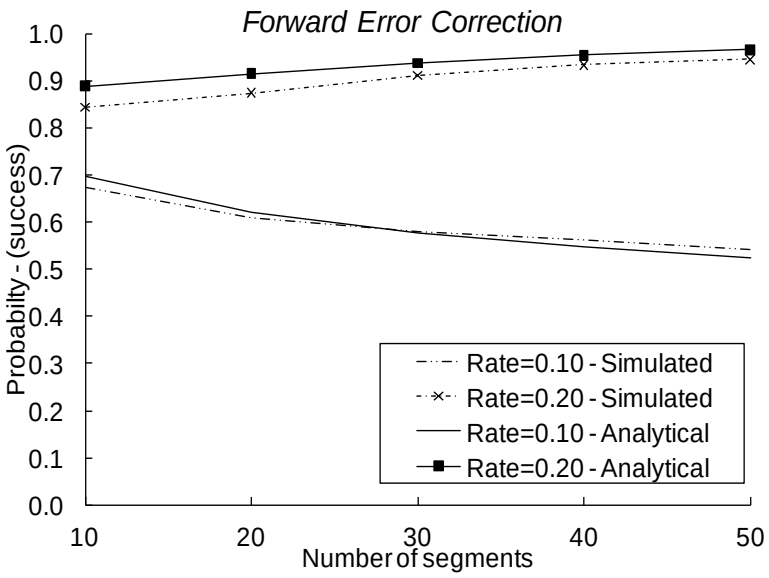


(b)

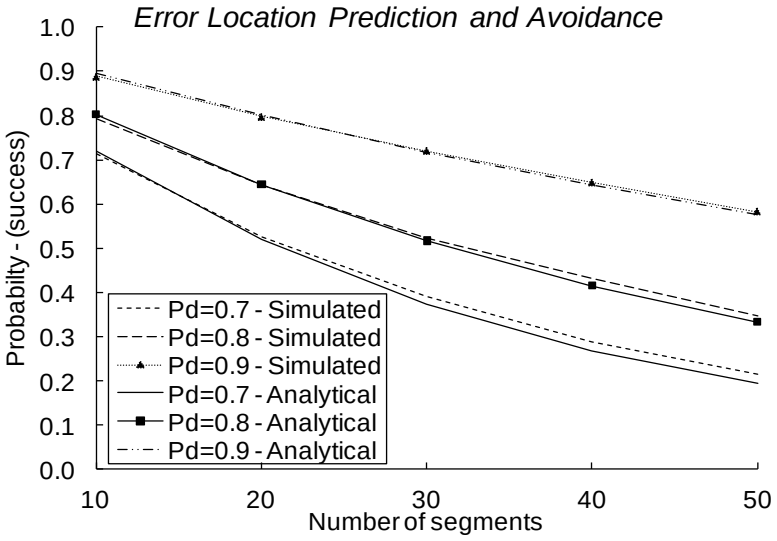
Figure 3.10: Simulated LMSC charatersitics probabilities(a) Probability that number of consecutive good state is more than given number of states (b) Probability that number of consecutive bad state is more than given number of states.



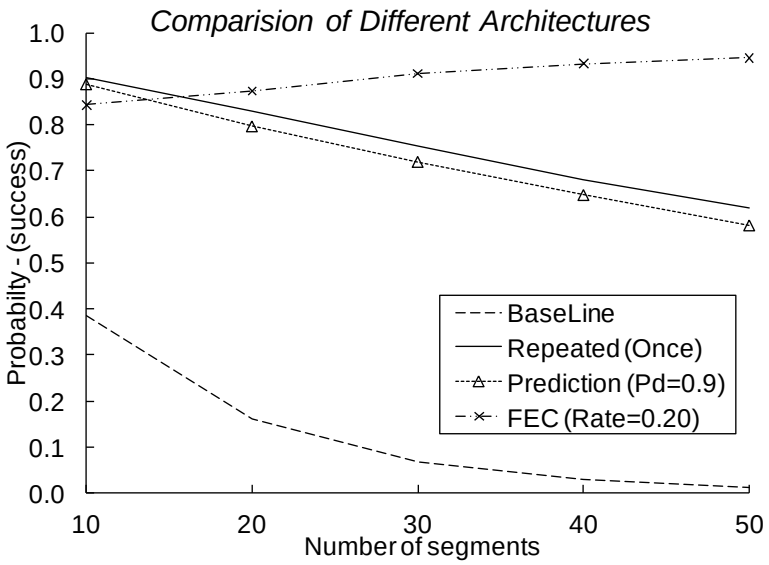
(a)


Forward Error Correction
Probabilty - (success)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10
20
30
40
50
Number of segments
Rate=0.10 - Simulated
Rate=0.20 - Simulated
Rate=0.10 - Analytical
Rate=0.20 - Analytical

(b)


Error Location Prediction and Avoidance
Probabilty - (success)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
10
20
30
40
50
Number of segments
Pd=0.7 - Simulated
Pd=0.8 - Simulated
Pd=0.9 - Simulated
Pd=0.7 - Analytical
Pd=0.8 - Analytical
Pd=0.9 - Analytical

(c)



(d)

Figure 3.11: Comparision of simulated and analytical success probabilities (a) BaseLine and Repeated transmission (n = 2) (b) Forward error correction (α = 0.1, 0.2) (c) Error location prediction and avoidance (pd = 0.07, 0.08, 0.09) (d) Comparision between simulated results of baseline, repeated transmission, forward error correction and error location prediction & avoidance architectures.

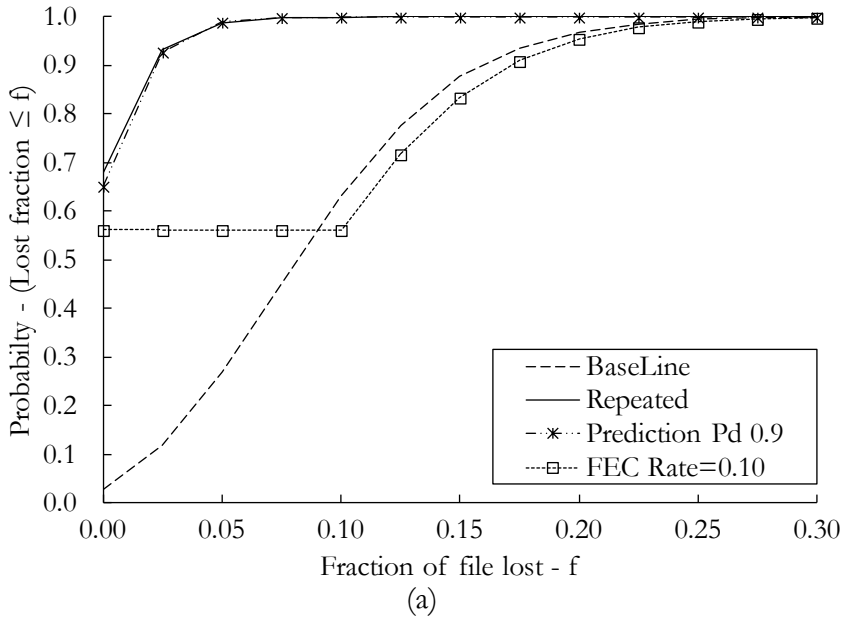
3.6 Conclusion

We have presented a viable solution for provision of the Internet access to the vehicular networks, especially during the initial deployment phase of vehicular networks and also in areas with very scarce roadside infrastructure (such as along highways and in rural areas). The solution is practical and economical since it only uses satellite receive-only terminals and very few (widely spaced) RSUs. We have also presented a number of error handling options which can be employed according to the operating environments. We have compared these options with mathematical analysis and simulation; both the comparisons agree with each other.

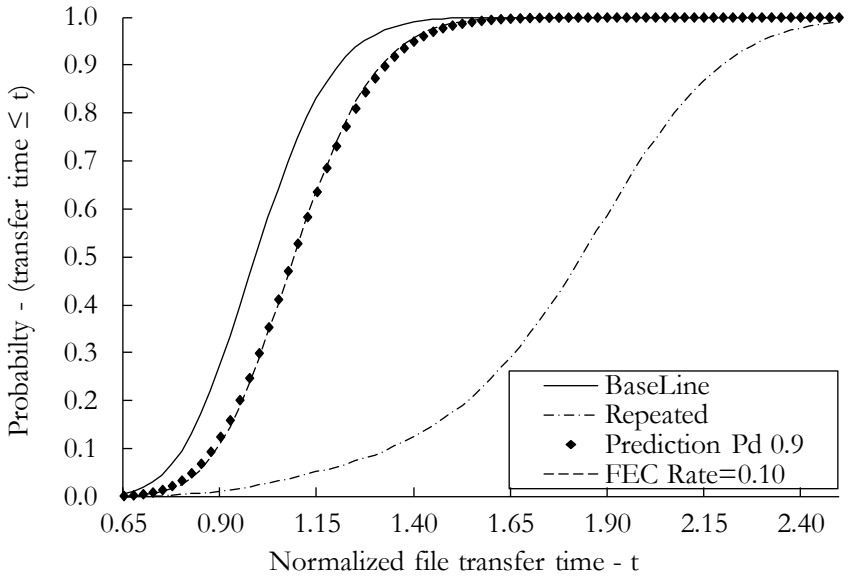
The efficiency of the solution can be further enhanced by using V2V communication in a variety of ways. For example, caching and later relaying the data for other vehicles (that might not have been

able to receive it due to error zone), relaying NAK to previous RSU (via vehicles traveling in opposite direction), using V2V communication as the reverse channel to send all the selective ACKs and NAKs, etc.

The solution is best suited for request-response type of applications, where a small request is followed by a large response data (such as file transfer, multimedia download, etc). The solution does not provide continuous connectivity so interactive or continuous connectivity demanding applications, such as IP telephony cannot be supported. Also, the solution is not intended to support security based applications that are time critical and require large data flow from vehicles; however, non-time critical or broadcast nature of security applications are supported, for example, dissemination of certificate revocation lists, weather, local news, hazard conditions, or other security alerts through satellites.



(a)



(b)

Figure 3.12: Comparision between baseline, repeated transmission (repeat once), forward error correction (Rate=0.10) and error location prediction & avoidance (P_d=0.9) architectures (a) Probability that the fraction *f* of file lost is ≤ given value (b) Probability that the normalized file transfer time *t* is ≤ given value. Note: transfer time has been normalized with mean transfer time of baseline architecture.

3.7 References

[1] M Bechler, WJ Franz, and L Wolf, "Mobile internet access in FleetNet", 13th Fachtagung Kommunikation in verteilten Systemen, April 2003.

[2] W Enkelmann, "FleetNet-applications for inter-vehicle communication", in IEEE Intelligent Vehicles Symposium, 2003.

[3] Franz, W. and Hartenstein, H. and Bochow, B., "Internet on the road via inter-vehicle communications", in GI/OCG Annual Conference: Workshop on Mobile Communications over Wireless LAN: Research and Applications, Sep 2001.

[4] J.A. Festag, H. Fußler, H. Hartenstein, A. Sarma1, and R. Schmitz, "FLEETNET: Bringing car-to-car communication into the real world", in 11th World Congress on ITS, October 2004

[5] J. Ott, and D. Kutscher, "Drive-thru Internet: IEEE 802.11b for automobile users", in IEEE Infocom Conf., 2004.

[6] Y. Yang, M. Marina, and R. Bagrodia, "Evaluation of multihop relaying for robust vehicular Internet access", in MObile Networking for Vehicular Environments (MOVE), 2007.

[7] V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, and Samuel Madden, "A measurement study of vehicular internet access using in situ Wi-Fi networks", in 12th annual intl. conf. on Mobile computing and networking, Sep 2006.

[8] R. A. Wyatt-Millington, R Sheriff, Y. F. Hu, P. Conforto, and G. Losquadro, "The SUITED project: a multi-segment system for broadband access to Internet services", in IEE Broadband Satellite Conf., 2000.

[9] J. Santa, R. T. Moreo, and A. F. G. Skarmeta, "A novel vehicle communication paradigm based on cellular networks for improving the safety in roads", in Int. J. Intelligent Information and Database Systems, Vol. 2, No. 2, pp. 240-257, 2008.

[10] D.G. Oh, P. Kim, Y.J. Song, I.J. Jeon, and H.-J. Lee, "Design considerations of satellite-based vehicular broadband networks", in IEEE Wireless Comm., vol. 12, Oct 2005, pp. 28-36.

[11] J. L. Mineweaser, J. S. Stadler, S. Tsao and M. Flanagan, "Improving TCP/IP performance for the land mobile satellite channel", in IEEE Military Comm. Conf., 2001

[12] C. Martin, A. Geurtz, and I. Ottersten, "File based mobile satellite broadcast systems: error rate computation and QoS based design", in IEEE 60th Vehicular Technology Conference, VTC2004-Fall, Vol. 6, pp. 4017- 4021, Sep 2004

[13] H. Ernst, L. Sartorello, and S. Scalise, "Transport layer coding for the land mobile satellite channel", in 59th IEEE Vehicular Technology Conference (VTC'04), May 2004.

[14] L. Casone, G. Ciccarese, M. D. Blasi, L. Patrono, and G. Tomasicchio, "An efficient ARQ protocol for a mobile geo-stationary satellite channel," in IEEE GLOBECOM, 2001, vol. 4, pp. 2692–2697.

[15] "IEEE Trial-Use Standard for Wireless Access in Vehicular Environments (WAVE) - Resource Manager," IEEE Std 1609.1-2006 , vol., no., pp.c1-63, 2006.

[16] "IEEE Trial-Use Standard for Wireless Access in Vehicular Environments - Security Services for Applications and Management Messages," IEEE Std 1609.2-2006 , vol., no., pp.0_1-105, 2006.

[17] "IEEE Standard for Wireless Access in Vehicular Environments (WAVE) - Networking Services," IEEE Std 1609.3-2010 (Revision of IEEE Std 1609.3-2007) , vol., no., pp.1-144, Dec. 30, 2010.

[18] "IEEE Draft Standard for Wireless Access in Vehicular Environments (WAVE) - Multi-Channel Operation," IEEE 1609.4/D8.0, June 2010 , vol., no., pp.1-92, July 15, 2010.

[19] "IEEE Draft Standard for Amendment to Standard [for] Information Technology-Telecommunications and information exchange between systems-Local and Metropolitan networks-

Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications-Amendment 6: Wireless Access in Vehicular Environments," IEEE Std P802.11p/D11.0 April 2010 , vol., no., pp.1-35, June 15, 2010.

[20] "IEEE Draft Standard for Information Technology-Telecommunications and information exchange between systems-Local and Metropolitan networks-Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications" IEEE Std P802.11-2007 March 2007 , vol., no., pp.1-1232, June 12, 2007.

[21] V. K. Garg, "Wireless Communication and Networking", San Francisco: Morgan Kaufmann, 2007

[22] ETSI EN 302 307 (DVB-S2); Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications.

[23] EN 301 790 (DVB-RCS); Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems.

[24] ETSI TR 102 768; Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790 in mobile scenarios.

[25] ETS 300 801 (DVB-RCP); Digital Video Broadcasting (DVB); Interaction channel through Public Switched Telecommunications Network (PSTN)/ Integrated Services Digital Networks (ISDN).

[26] ETSI EN 301 958 (DVB-RCT); Digital Video Broadcasting (DVB); Interaction channel for Digital Terrestrial Television (RCT) incorporating Multiple Access OFDM.

[27] T. T. Kwan , R. E. McGrath , D. A. Reed, "User access patterns to NCSA's World Wide Web server", University of Illinois at Urbana-Champaign, Champaign, IL, 1995.

[28] H. Clausen, H. Linder, and B. Collini-Nocker, "Internet over direct broadcast satellites", Communications Magazine, IEEE, vol. 37, 1999, pp. 146-151

[29] A. Jamalipour, "The Wireless Mobile Internet - Architectures, Protocols and Services", Wiley, 2003

[30] J. S. Baras, S. Corson, S. Papademetriou, I. Secka, and N. Suphasindhu, "Fast asymmetric Internet over wireless satellite-terrestrial networks," in MILCOM '97, Nov. 1997, pp. 372-377

[31] G. Fairhurst, N.K.G. Samaraweera, M. Sooriyabanadara, H. Harun, K. Hodson, and R. Donadio, "Performance issues in asymmetric TCP service provision using broadband satellite", in IEE Comm, Vol 148 No 2, April 2001.

[32] I. Minei and R. Cohen, "High-speed Internal access through unidirectional geostationary satellite channels, in IEEE JSAC, vol. 17, no. 2, pp. 345-359, Feb. 1999.

[33] J. Ott, and D. Kutscher, "The drive-thru architecture: WLAN-based Internet access on the road", in Vehicular Technology Conference VTC 2004-Spring, 2004.

[34] iCartel: MIT CarTel. http://icartel.net/icartel-docs/

[35] J. Gutiérrez. "Selected readings on telecommunications and networking", Idea Group Inc (IGI), 2008.

[36] Data Calculator: http://www.att.com/standalone/data-calculator.

[37] J. Luo and J.-P. Hubaux, "A survey of research in inter-vehicle communications", in Securing Current and Future Automotive IT Applications, pp 111-122, Springer-Verlag, 2005.

[38] K Levacher, F McGee, F Murphy,"A comparison between 3G and 802.11 wireless technologies for Inter-Vehicular Communications purposes", http://killian.levacher.googlepages.com/Acomparisonbetween3Gand802.11wireles.pdf

[39] Y. F. Ko, M. L. Sim, and M. Nekovee, "Wi-Fi based broadband wireless access for users on the road", BT Technology Journal, vol. 24, pp. 122- 129, April 2006

[40] A. Qureshi and J. Guttag, "Horde: separating network striping policy from mechanism", in 3rd intl. conf. on Mobile systems, applications, and services, June 2005.

[41] R. Chakravorty, A. Clark and I. Pratt, "GPRSWeb: optimizing the web for GPRS Links", in ACM/USENIX MobiSys, San Francisco, May 2003.

[42] M. C. Chan and R. Ramjee, "TCP/IP performance over 3G wireless links with rate and delay variation", in ACM Mobicom, Sep 2002.

[43] G. Giambene and D. Miorandi, "A simulation study of scalable TCP and highSpeed TCP in geostationary satellite networks", in J. Telecomm systems, Vol 30, No 4, pp. 297-320, Dec 2005.

[44] D. Barman, I. Matta, E. Altman, and R. ElAzouzi, "TCP Optimization through FEC, ARQ and Transmission Power Tradeoffs," in Wired/Wireless Internet Comm. conf., Feb. 2004.

[45] N. Celandroni, "Comparison of FEC types with regards to the efficiency of TCP connections over AWGN satellite channels", IEEE Transactions on wireless communications, Vol 5, No 7, pp. 1735-1745, July 2006

[46] P. Papadimitriou, and V. Tsaoussidis, "On TCP performance over asymmetric satellite links with real-time constraints", in J. Computer Comm., Vol 30 No 7, pp. 1451-1465, May 2007.

[47] H. Obata, K. Ishida, J. Funaska, and K. Amano, "TCP Performance Analysis on Asymmetric Networks Composed of satellite and terrestrial Links", in. IEEE Intl. Conf. on Network Protocols, ICNP'2000, pp. 199-206, 2000

[48] G. Giambene and M. Marandola, "Internet access in hybrid terrestrial and satellite mobile communication systems", in IEEE VTC, 2004.

[49] T. A. Gillespie, "Modeling the transformational communications system urban land mobile satellite channel", 2007, Storming Media, http://www.stormingmedia.us/14/1406/A140674.pdf

[50] D. Roddy, "Satellite Communications", (eBook) 4Th Ed., New York: McGraw-Hill Professional, 2006.

[51] Giovanni E. Corazza, "Digital Satellite Communications". New York: Springer, 2007.

[52] E. Lutz, M. Werner, and A. Jahn, "Satellite systems for personal and broadband communications", Berlin, Germany: Springer-Verlag, 2000.

[53] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, "The land mobile satellite communication channel—Recording, statistics and channel model", IEEE Trans. Veh. Technol., vol. 40, no. 2, pp. 375–386, May 1991.

[54] S. Scalise, H. Ernst, and G. Harles, "Measurement and modeling of the land mobile satellite channel at Ku-band," in IEEE Trans. Veh. Tech., Vol 57, No 2, pp. 693-703, March 2008.

[55] C. Martin, A. Geurtz, and B. Ottersten, "Statistical analysis and optimal design for efficient mobile satellite broadcast with diversity", in IEEE Trans. on Veh. Tech., Vol 57, No 2, pp 986-1000, March 2008.

[56] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow, "TCP selective acknowledgement options", IETF RFC 2018, October 1996.

[57] L. Eggert and F. Gont, "TCP user timeout option", standards track (draft-ietf-tcpm-tcp-uto-11), January 22, 2009.

[58] J. Border, M. Kojo, J. Griner, G. Montenegro, and Z. Shelby, "RFC 3135: Performance enhancing proxies intended to mitigate link-related degradations", IETF RFC 3135, June 2001.

[59] L. Rizzo, "Effective erasure codes for reliable computer communication protocols," ACM Computer Communication Review, April 1997.

[60] V. Jacobson, "Congestion Avoidance and Control," in ACM SIGCOM'88 Conference, pp. 314-329, August 1988.

CHAPTER 4 SECURITY ARCHITECTURE

The desired security attributes for VANET include authentication, confidentiality, integrity, non-repudiation, revocation and privacy. It is important to note that privacy is the most important attribute, but at the same time it is in conflict with other attributes thus complicating the design of VANET security architecture.

The simplest security architecture is to assign a single permanent certificate to each vehicle, this ensures authentication, confidentiality, integrity, non-repudiation, revocation but not the privacy. To address privacy, basic architecture can be extended to use multiple temporary certificates (normally referred as pseudonyms) instead of one permanent certificate; this ensures privacy since pseudonyms cannot be linked with each other and to the user [1-6]. Different schemes for pseudonym-management have been proposed to ensure unlink-ability. One such scheme is to issue pseudonyms in bulk to vehicles [1]; the vehicle can then use these to ensure privacy. The bulk pseudonyms based scheme requires a tamper-proof-device (TPD) to store the pseudonyms and perform cryptographic operations [1], since these pseudonyms may be used for malicious purposes such as Sybil attacks. The TPDs are expensive and need reloading with new pseudonyms when old ones expire or are used up.

Possible solutions can be to let vehicles generate pseudonyms themselves [2, 3] or periodically get new pseudonyms from some certificate servers [4, 5]; thus eliminating the need of TPD (given the pseudonyms/other-authenticating-credentials with overlapping validity are not generated in bulk). First option makes revocation very complex and difficult while second option makes privacy difficult to achieve (since certificate server can link various pseudonyms). Blind signature scheme [7], with some kind of link-ability, is usually employed to address privacy issues of second option [4, 5]. The process requires multiple-certificate-servers/multiple-transactions for one signature (i.e., for getting one pseudonym) and is thus difficult to realize, especially with an intermittent communication link with the infrastructure. Blind signature scheme is also used in [6], but the solution requires generation of authenticating-tokens in bulk thus needing TPD.

Other architectures include those based on principles of group signatures and ID cryptography [8]. In case of group signatures, vehicles form part of a group with a trusted group manager. The architecture requires members to trust the group manager (who can find the true identity of signer), which will be difficult to achieve in a dynamic VANET. Further, size, membership revocation and dynamic membership (new nodes entering a group and old nodes leaving the group) increase the complexity and overheads of this method.

The centralized certificate authority (CA) based solutions present a number of challenges which may be difficult to address during the initial deployment stages of VANET. The CAs must be organized in a hierarchical manner for effective management. The hierarchy can be area/location based; a given area (e.g., United States or Europe) can be divided into regions (e.g., states or countries) with

each region having its regional CA, these regional CAs are then linked with each other via a top level CA. Figure 4.1 shows a hierarchy with two regions.

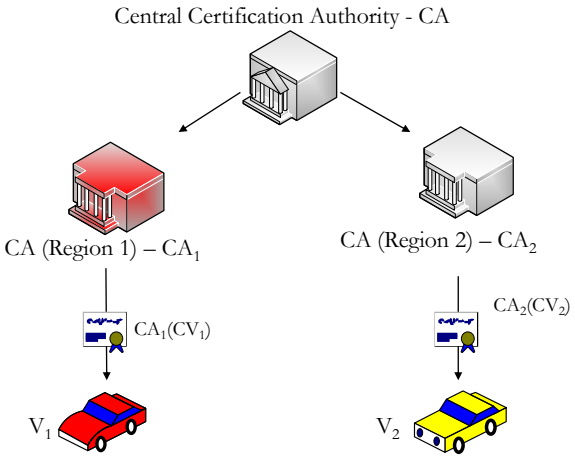


Figure 4.1: A certification authority hierarchy with two regional CAs. CA (Region 1) issues certificates to vehicles registered with in its region, for example certificate CV_1 is issued to vehicle V_1. (Note: $CA_x(CV_y)$ is a certificate issued to vehicle y by a CA of region x.)

The hierarchy can be extended both upwards and downwards. This means for vehicles to easily travel outside their CA's domain, we need to establish a trust relationship among all certification authorities; thus certificate verification may take longer if the trust relationship goes through a long chain. Figure 4.2 shows possible steps taken for certificate verification when a vehicle from one region tries to communicate with a vehicle from another region (it assumed that none of the intermediate entities have previously cached certificates).

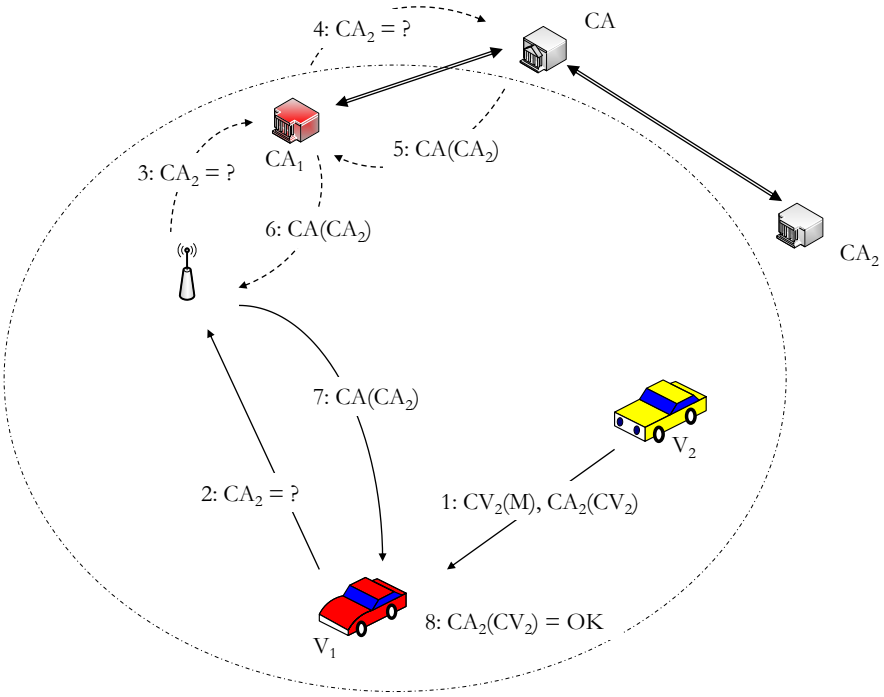


Figure 4.2: Certificate verification. (1) V_2 sends a signed message along with its certificate to V_1. V_1 does not have certificate CA_2 in its cache and therefore cannot verify CV_2. (2, 3) V_1 asks for CA_2 from its regional CA via roadside unit. (4) Regional CA may have to ask central CA for the CA_2. (5, 6, 7) Certificate $CA(CA_2)$ is sent to V_1 via regional CA and roadside unit. (8) V_1 verifies the certificate CA_2/CV_2 and accepts the message. (Note: $CA_x(CV_y)$ is a certificate issued to vehicle y by a CA of region x and dotted circle indicates a region.)

Further, it also makes revocation difficult since revocation list (RL) must be distributed to all regions as vehicles are not restricted to remain within their regions. Figure 4.3 shows the distribution of RL in case of two regional CAs. If pseudonyms are preloaded in TPDs then certificate revocation for a particular vehicle must include all the pseudonyms currently issued to (stored in) the vehicle. The RL may grow over time, making its distribution more difficult.

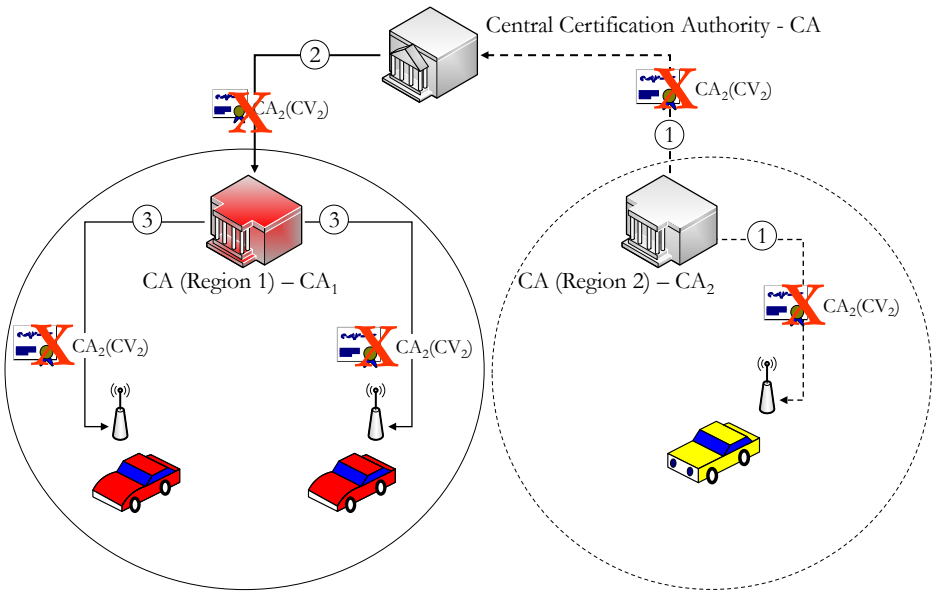


Figure 4.3: Distribution of certificate revocation list. (1) CA (Region 2) revokes certificate of a vehicle in its region, it distributes the revocation information within its region and also forwards it to central CA. (2) Central CA forwards revocation information to all regional CAs. (3) Each regional CA disseminates revocation information within its region. (Note: $CA_x(CV_y)$ is a certificate issued to vehicle y by a CA of region x and circles indicate regions.)

Each vehicle will have an associated certificate since its manufacture, this will be modified or updated each time the owner changes. These certificates will be expensive and it will also be technically difficult for an average user to keep track of the certificate renewal etc (even if he is not using the services). Further, in case of possible compromise, the revocation and issuance of new certificate may be quite cumbersome.

In current designs, too much trust is placed on TPD, which stores all cryptographic materials (permanent certificate and pseudonyms), performs cryptographic operations (signing/verifying messages) and processes revocation messages/commands (erase keys/pseudonyms when revoked) [10]. Since the vehicle (and TPD) cannot be physically guarded as other electronic security devices

(smart cards etc), those requirements will make the device quite expensive [14]. Further, the pseudonyms when exhausted must be reloaded thus requiring a periodic maintenance.

The initial deployment stage of VANET will be characterized by limited infrastructure and small number of smart vehicles, which means very limited vehicle to vehicle and vehicle to infrastructure communication. During this stage, the solutions that assume omnipresence of these communications for certificate issuance, verification or revocation will not be practicable. Further, lack of infrastructure will discourage consumers' participation and lack of consumers (smart vehicles) will discourage providers' investment in infrastructure.

In order to achieve the desired security attributes, two different distributed security architecture for VANET that do not rest on expensive security hardware or elaborate security infrastructure. are presented:

- **Service oriented security architecture**: The architecture is based on spatial and temporal restricted certificates, which are issued upon user's request and can be used for various VANET applications. Due to the restricted nature of these certificates, the certificate revocation process is simple and efficient. The architecture can be incrementally deployed, facilitating small companies to jump in the VANET business, and can fill the void during the VANET initial deployment phase.
- **General purpose security architecture**: The security architecture uses revised Blind signature scheme. It provides "one-way-link-ability" that helps to achieve all the security attributes without introducing complex/multi-transaction procedures. It does not require expensive TPDs or complex pseudonym issuance/revocation procedures and is especially

suited to VANET during initial deployment phase which is characterized with intermittent connectivity. Further, non-repudiation/revocation requires cooperation between multiple entities thus ensuring privacy without a single point of failure.

The chapter is organized in 7 sections. Section 4.1 discusses related research work in the field. Section 4.2 introduces the system model. Section 4.3 discusses the proposed service oriented security architecture. Section 4.4 discusses the proposed general purpose security architecture. And in the end, section 4.5 presents conclusion.

4.1 Related Work

Papadimitratos et al. [10, 11, 1] have presented a quite comprehensive solution based on central/regional certification authorities and their trust relationships. The solution uses pseudonyms to address privacy issues. The pseudonyms are preloaded in TPD [10] or issued by pseudonym provider [11] or generated by TPD and signed by CA [1]. They have also highlighted multiple revocation protocols. The solution requires the TPD, of the vehicle whose certificates have been revoked, to delete all stored pseudonyms and also assumes CA to have some knowledge about vehicles location. A malicious node may avoid this deletion by blocking the revocation message. This may enable him to use the pseudonyms later for communication with other vehicles. Other options are distribution of compressed RL or using bloom filters. TPD management through signed messages from CA may be exploited to evade revocation or for other malicious purposes such as DoS attacks (causing victim's TPD to delete key material, etc). [1] leaves misbehavior detection on vehicle between infrequent RL distributions.

The distribution of RL to all smart vehicle/regions is also a challenge. Papadimitratos et al suggest restricting the scope of RL within a region, and requiring visiting nodes from other regions to obtain temporary certificates [15]. Thus a vehicle will have to acquire temporary certificates if it is travelling outside its registered region.

In [2] Armknecht et al. propose a public key infrastructure where users derive public keys, certificates and pseudonyms. The architecture is based on elliptic curves, each user gets a master key and master certificate from CA. It can then generate its key pairs or certificate using masker key, master certificate and its own secret key. The certificate generated by user is verifiable by CA's public key. For revocation the CA publishes some data depending on which all nodes have to update their keys. The excluded nodes cannot update the keys based on this data. This means for each revocation everybody has to update their certificates.

In [3] Fan et al. present detailed operation of public key infrastructure mechanism based on bilinear mapping. They achieve privacy through pseudonyms which are generated by users themselves similar to [2]. Revocation is accomplished through distribution of RL that is stored by each user. Every time a user receives a beacon it performs certain computations on complete RL to ensure that the received beacon is from unrevoked user.

In [4] Rahman et al. present an automated crash reporting application. For privacy, they use Blind signature scheme to get anonymous credentials signed by local certification authority (government transportation authority -GTA) through a multiple transaction protocol. They achieve non-repudiation by adding an invisible identity field in pseudonym. A vehicle's unique identity (within a

GTA's domain) is doubly encrypted (first by GTA's public key then by local law enforcement authority's public key) to get an invisible identity. They suggest using cut-and-choose method to ensure that blind messages are well formed, which has high overheads especially to confirm the invisible-identity. Further, the cut-and-choose method will reveal the identity of vehicle thus compromising privacy.

In [8] Lin et al. present a security mechanism using group signature and identity based signature techniques. The solution minimizes the storage at CA for later liability establishment, however the revocation is road side unit aided. CA sends RL to roadside unit which then monitors certificates in messages broadcasted by passing-by vehicles and if a message with revoked certificate is observed then roadside unit broadcasts warning messages. In another option it is suggested that each passing-by vehicle get its certificate signed from roadside unit. These signatures are then used to show that the certificate has not been revoked. First option is open to attacks (malicious node does not transmit within range of a roadside unit) and second increases complexity and overhead.

Our general purpose security architecuture comes closer to the method presented by [5, 6]. In [5], Fisher et al. used a large number of pseudonyms (defined as Inter-Vehicle-Communication-IVC certificates) to achieve un-link-ability. These pseudonyms are blindly signed by IVC certification servers' (ICS) private key. The private signing key is shared amongst multiple ICS by means of Secret Sharing. An IVC certificate is distributedly calculated through a quorum of ICS. For non-repudiation a tag, that can be linked to the vehicle, is generated/stored by ICS and is protected by a secret key shared amongst ICS. The solution requires transactions with multiple servers to get a pseudonym which may be difficult due to intermittent connectivity in VANET. Further, a

pseudonym cannot be revoked during its validity period, and no definite solution to malformed pseudonyms (having validity larger than defined maximum period) has been defined. In [6], Schaub et al. also use pseudonyms to achieve un-link-ability. The pseudonyms are issued by pseudonym providers (PP_k) based on V-tokens (that also later form part of pseudonyms), V-tokens cannot be linked to the each other or to the owner by PP_k thus ensuring privacy. V-tokens, containing identifying information of the vehicle and Certification Authority (CA), are blindly signed by CA after being encrypted by vehicle with public key of resolution authority (RA). The decryption ability of V-tokens (i.e, resolution/non-repudiation) is distributed using threshold encryption scheme. The solution relies on cut-and-choose method to ensure well-form-ness of V-tokens, thus adding overheads in addition to the need of TPD (to store the V-tokens or corresponding pseudonyms). Further, the revocation method only revokes long-term identity and does not address already issued pseudonyms/V-tokens which may continue to be used for malicious purpose.

IEEE P1609.2 [9] proposes a CA based architecture. The architecture assumes pervasive roadside architecture and also does not offer certificate revocation options.

In [16] Parno et al. present detailed discussion on challenges faced by vehicular network, adversaries, attacks and propose a set of security primitives. They suggest a dynamic key distribution system, where each node generates its own short term key pair and requests CA to issue a certificate based on generated public key. They also suggest using group signatures to achieve anonymity.

In [17] Lin et al. present a security mechanism using group signature and identity based signature techniques. The solution minimizes the storage at CA for later liability establishment, however the

revocation is road side unit aided. CA sends RL to roadside unit which then monitors certificates in messages broadcasted by passing-by vehicles and if a message with revoked certificate is observed then roadside unit broadcasts warning messages. In another option it is suggested that each passing-by vehicle get its certificate signed from roadside unit. These signatures are then used to show that the certificate has not been revoked. First option is open to attacks (malicious node does not transmit within range of a roadside unit) and second increases complexity and overhead.

4.2 System Model

4.2.1 Security Objectives

VANET's security requirements are more complex than other wired/wireless networks. In addition to basic security attributes of authentication, confidentiality and integrity, it also requires non-repudiation, revocation and privacy. These additional security attributes are briefly discussed below:-

- **Non-repudiation**: A user should not be able to later deny that she originated a message. It adds liability to user for the messages which she generates. This is especially important in case of VANET safety applications. If this requirement is not fulfilled then a malicious node may generate fake public safety message without any liability.
- **Revocation**: Revocation of user's credentials is also an important security attribute. It helps to minimize the damages if a user's credentials are lost or a user engages in malicious activity.
- **Privacy**: Privacy is one of the most important security attributes in VANET applications. This is due to the fact that VANET communication can be used to track a vehicle (driver) which causes great concerns to many users. Privacy comes in direct conflict with the other

security attributes. One has to strike a balance between privacy protection and the other security attributes, especially non-repudiation.

4.2.2 Threat Model

We do not make very stringent security requirements for vehicle's on-board device or restrict the capabilities of attacker node. We assume that an attacker is capable of:

- eavesdropping when within the routing path or in the transmission range of a message
- injecting, modifying, spoofing or dropping the messages
- trying to track the movement of another vehicle either alone or in collaboration with other mobile or fixed nodes (total number of such collaborating nodes will be a small fraction of all the nodes participating in the network since we assume that majority of nodes are honest)
- taking complete control of her on-board device and also crafting any protocol related messages

4.2.3 Desired Requirements

Keeping in mind VANET characteristics, attacker capabilities and security attributes, our desired requirements for the proposed security architecture are:-

- Ensure authentication, confidentiality, integrity, non-repudiation, revocation and privacy.
- Guard against traceability by one or more collaborating entities. An attacker alone or with collaboration of limited other mobile or fixed nodes should not able to track a user. In other words, two messages from the same user should not be linkable (if desired).

- Ensure privacy revocation involves multiple authorities. A single authority, by itself alone, should not be able to revoke the privacy of a user. Privacy revocation could only be achieved by cooperation of multiple identities.
- Provide security without need of expensive TPDs, or large storage requirements at central authority/ RSU.
- Guard against a user using legitimate pseudonyms for malicious purposes such as Sybil attack, etc.
- Do not require multiple transactions for various routine operations, such as certificate issuance, certificate revocation, etc. This is especially necessary due to the intermittent nature of connectivity of VANET.

4.3 Service Oriented Security Architecture

To address the security challenges during initial deployment stage, a distributed certificate architecture is proposed. This stage will be characterized by very few smart vehicles and lack of necessary roadside infrastructure to support various VANET applications or elaborate security architecture. The proposed architecture achieves desired security attributes and enables service providers to offer incrementally various VANET services with minimal investment thus encouraging both service providers and users to try/adopt VANET. Certificates with a limited scope in both time and space domain are issued by a service provider. These certificates are usable within a particular geographic area or within a certain time or both. These certificates are not tied to the vehicle's registration etc and can be changed periodically during one service period. Meanwhile law enforcement agencies can trace back the user via the temporary certificate and the service provider.

4.3.1 Assumptions

Our solution is based on a few simple assumptions given below:

- The user/node (we use user/node/vehicle interchangeably in this paper) has a payment-processing-device (similar to automatic toll payment devices - sold for tens of dollars). We do not require the device to store pseudonyms, perform cryptographic operations (such as signing/verifying messages) or perform revocation operation. The device only participates in credential/service request operations (discussed later).
- The user/node has a wireless-communication/VANET-application device that can communicate with roadside infrastructure; it can be a laptop or a hand-held device or a device specially designed for smart vehicles. The device can communicate (wired/wireless/WiFi/Bluetooth) with the payment-processing-device.
- Limited local roadside units are available (the existing hotspots in urban areas may be used for this purpose) and service providers can be accessed through these roadside units.

4.3.2 Basic Solution

The basic solution only caters for the provision of temporary credentials so that the required security attributes are achieved. These temporary credentials (pseudonyms) can then be used for basic vehicle to vehicle communication or participation in VANET safety application (such as initiating/relaying safety information).

The basic idea is that if a user wants to participate in a VANET (the user's vehicle is not required to have a manufacturer's issued certificate), he purchases a payment-processing-device (As mentioned above, it is assumed that user also has a VANET application device, which is running desired VANET applications). Each device will have an identification and an associated certificate. During initialization the device will be linked/registered with the user's account. The user's information will be maintained with the provider and will not be stored in the device. The basic procedure is illustrated in Figure 4.4. When a user enters a service area and wants to use the service, he makes the payment for the service using onboard payment device. The payment-authorization/service-request message will be encrypted using the provider's public key, thus hiding the device ID/certificate and services requested from eavesdroppers. The user is issued a pseudonym by the provider that will be valid for a given period/area.

We define several notations/functions that we will use in the formal description of our solution. A certificate or a pseudonym will essentially be represented by its public and private key pair; such as (K_x^+, K_x^-) are public (+) and private (-) keys belonging to *X*. (t_s, t_f) are the start and finish times between which a particular pseudonym (*P*) will be valid. A certificate can be valid inside a service area; service areas can be defined with region numbers *R,* large service areas may have more than one region. A user specifies the region and time period, in the request, for which he/she wants to purchase the certificate. $E_{K^+}(M)$ defines an encryption function on message *M* using the public key K^+. Public cryptography is very resource intensive therefore data encryption is usually carried out using a randomly generated symmetric session key and only the session key is encrypted using public cryptography. The encryption function $E_{K^+}(M)$ defined above employs similar techniques; we will not show the details for simplicity and compactness. $S_{K^-}(M) = N$, defines a signature function on

message M using a private key K^-. The signatures are computed by first creating a message digest using a hashing function and then encrypting the digest using key K^-. $V_{K^+}(M,N)$ is a signature verification function. It has two inputs the message M and the signature N. It verifies the signature by computing the message digest of message M and comparing it with received signatures N (after decrypting it with the corresponding public key K^+).

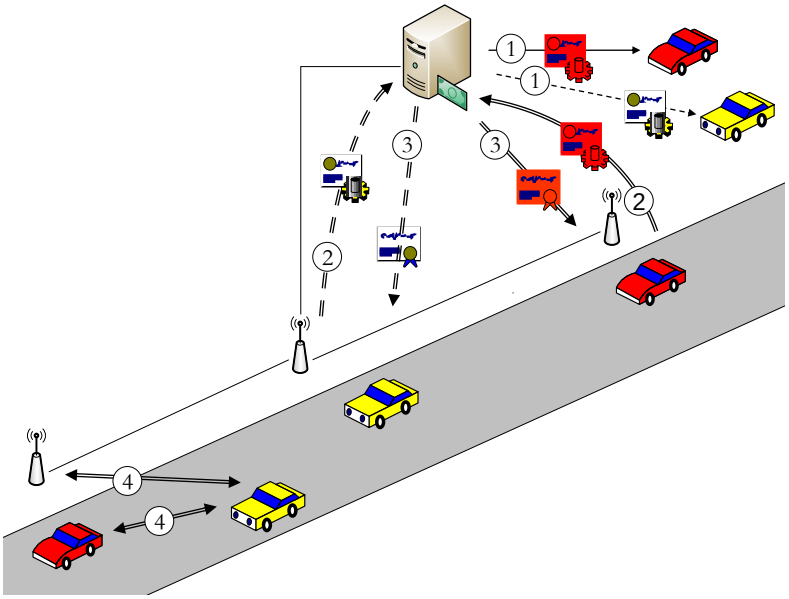


Figure 4.4: Architecture (1) Users register their payment devices with Provider beforehand (2) Users send payment/service requests (3) Provider issues temporary credentials (4) Users participate in VANET via vehicle to vehicle or vehicle to infrastructure communication.

If a user U having a public key pair (K_U^+, K_U^-) (for initial request these are the permanent keys associated with the payment-processing device) wants to acquire temporary credentials for the time duration defined by (t_s, t_f) and within the region R from a service provider S with a public key pair (K_S^+, K_S^-), Figure 4.5 shows the transactions.

1	U :	Generate $M = \{t_s, t_f, R, U\}$
		Compute $M_R = E_{K_s^+}(M)$
		Compute $N_R = S_{K_U^-}(M)$
2	$U \rightarrow S$:	M_R, N_R
3	S :	Extract M
		$V_{K_U^+}(M, N_R)$ [a]
		Verify ID U and associated account
		Generate $P = \{t_s, t_f, R, K_P^+\}, K_P^-$ [b]
		Compute $M_P = E_{K_U^+}(P, K_P^-)$
		Compute $N_P = S_{K_S^-}(P)$
		Compute $N_K = S_{K_S^-}(K_P^-)$
4	$S \rightarrow U$:	M_P, N_P, N_K
5	S :	Extract P and K_P^-
		$V_{K_S^+}(P, N_P)$, $V_{K_S^+}(K_P^-, N_K)$

[a] The service provider records device's public key during user/device registration/initialization process.
[b] P is the pseudonym/temporary certificate with associated private key K_P^-

Figure 4.5: Transactions between User *U* and Provider *S* to acquire temporary credential $\{t_s, t_f, R, K_P^-, K_P^+\}$; valid for time duration defined by (t_s, t_f) and within region *R*. User uses *(P, N_P)* as a temporary certificate.

4.3.3 Extended Services

The solution can be easily extended for extended/additional services. If additional VANET services or applications are available (such as multimedia content, web access, email etc) then these can be offered as extended services. In this case a user indicates the service which he desires to use/purchase in service request/payment authorization message. The payment processing provider issues the temporary credentials to the user and also forwards these credentials along with the details of service purchased to the concerned server. The user can then initiate request to the concerned

server for service using issued temporary credentials. Figure 4.6 shows such a scenario. The Extended services will include the basic service (basic service only provides pseudonym).

It is not necessary that the payment-processing provider is also operating the application servers; these servers can be operated by other providers. In this case, the payment-processing provider provides temporary credentials and processes the payments on behalf of other providers; similar to credit card providers.

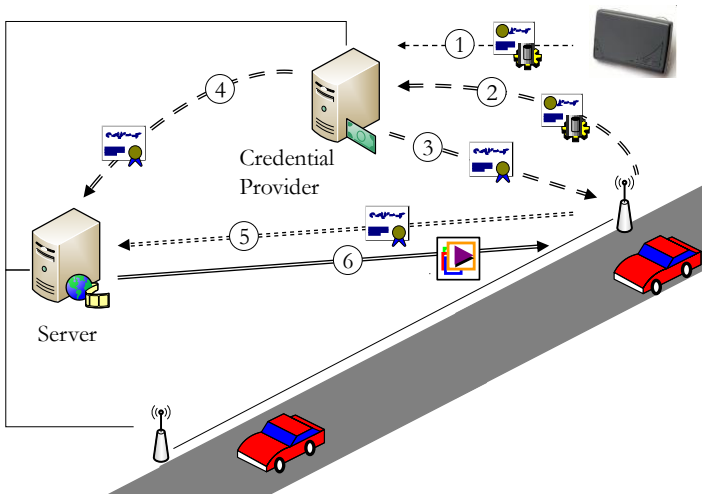


Figure 4.6: Extended services architecture (1) User registers payment device with Credential provider (2) User sends payment/service request (3) Credential provider issues temporary credentials (4) Credential provider informs Server of service purchased and temporary credentials (5) User requests service using temporary credentials (6) Server delivers content.

4.3.4 Provision of Privacy

The privacy is one of the most important security attributes in VANET. The proposed solution provides this through pseudonyms which cannot be linked to the user ID. For additional privacy the

pseudonyms can be refreshed within one service period. There are two possible options for this; the pseudonyms are issued in bulk at the time of purchase or a new pseudonym is issued sometime before the expiry of the old pseudonym.

In case of bulk issuance of pseudonyms there are a few aspects to be considered. The number of pseudonyms is related to time period for which the service has been purchased and desired level of privacy (i.e. how often the pseudonyms are changed). (In this paper we are not considering the exact time period or methodology for changing pseudonyms; this has been studied in detail by other researchers [1, 18-20, 23].). The validity period of each pseudonym is also important. If multiple pseudonyms have overlapping validity periods, they may be used for Sybil attacks. Although each pseudonym can be traced back to the user via a payment-processing-provider, this can only be done by law enforcement/government agencies and not by ordinary users. Another important aspect is the length and number of messages that are required to send these pseudonyms to the user/server and also the storage requirement at server/user device. If the pseudonyms are sent in one or multiple continuous messages, a malicious server (not the credential provider) may be able to link the pseudonyms and compromise user's privacy. For this reason, the credential provider should first mix/group the pseudonyms of different users (that will be served by same server) with each other and then send them to a service provider. User's applications also need to be careful about changing the pseudonyms to ensure security and uninterrupted service, for example not changing a pseudonym within a transaction or between multiple transactions that can be linked based on context (accessing one's email).

In case of single issuance of pseudonym the most important aspect is to ensure that the user gets new pseudonym before the expiry of current one. There are two options for this, either the user initiates request for a new pseudonym before the expiry of current one or the server maintains state for each user and issues a new pseudonym before the expiry of current pseudonym. Letting users initiate requests is more practicable since it will save server's resources and the complexity of message delivery (the user can initiate request anywhere within the service area).

Besides certificates (pseudonym) other IDs (such as IP address, MAC address etc) are also important to hide in ensuring privacy [21, 22]. These IDs can be issued on temporary basis and refreshed several times during a service period similar to pseudonyms.

The certificate of CA (also the payment-processing-provider) is hard coded in the payment device, enabling other users to check the validity of a certificate.

4.3.5 Practicability

The proposed solution is incremental, practicable and requires minimal infrastructure, which is especially advantageous during the initial deployment phase of VANET. The payment-processing-device does not need to have many functionalities or high processing power or large storage. It is similar to toll-payment-devices which are commonly being used and can be purchased for tens of dollars.

The payment-processing-device is not tied to a particular vehicle so a user is free to transfer it from one vehicle to another. The payment-processing provider is similar to credit card providers; we are using the mature Internet-like payment-processing architecture which is considered to be secure.

The application servers can be installed by different operators and existing hot spots in urban areas may be initially used to test the architecture.

Software can be developed for laptops and handheld devices to participate in different VANET applications. This will also provide a framework where different VANET applications can be tried or tested.

4.3.6 Analysis

The proposed architecture ensures desired security attributes. Authentication and confidentiality can be achieved by signing/encrypting the messages using associated public keys. Attacker cannot link the pseudonyms with a user; even different pseudonyms cannot be linked with each other, thus ensuring privacy. Meanwhile, liability can be enforced with the help of payment processing provider, since it has the account information for each issued pseudonym.

The architecture, as opposed to existing solutions, does not require users to maintain permanent (long-term) or valid temporary certificates when they are not using the service; user purchases a certificate only when he wants to use the service. The architecture also simplifies the certificate

revocation; certificates automatically expire after their validation time or beyond the predefined service area. For each new issuance of a certificate the provider checks if a previous certificate for the same user was revoked (each user account has an associated revocation flag that indicates whether a previous certificate of user was revoked or not. The provider can reset the flag if the user later clears the cause of revocation). If a revocation entry exists then new certificate will not be issued. Further, if the certificate is to be revoked before its expiry then revocation list (RL) can be disseminated via roadside units. Since the service is area/time restricted so the RL will be distributed only within the affected area and will contain only the certificates which have still not expired (due to time). This simplifies RL maintenance and distribution.

The system does not require centralized CA or trust relationships among regional CAs. Each provider can work independently within its coverage area. This minimizes the infrastructure required by a service provider to start its services and will be an incentive for service providers and facilitate small companies jumping into the VANET industry. Initially, a service provider may limit its service within a geographic area and later incrementally extend it. Further, when isolated/widely-separated service areas become adjacent due to the extensions then they can be combined as one region or roaming can be coordinated between the regions. Users and providers both benefit with incremental deployment without paying unnecessarily for the services they do not use or sell. The solution does not require expensive tamper proof devices and periodic refilling of pseudonyms. A user only pays when using the service and does not pay for certificate maintenance.

Payment devices may be operated by a third party and integrated with service providers; one device may be used by different service providers. Further, development of payment device will be

motivated by service providers, who will force security and affordability of the devices. The architecture derives its security from the mature Internet payment systems.

As a baseline service, the temporary credentials can be used for all VANET applications including vehicle to vehicle communications. Further, our solution can coexist with the solutions that are based on the certificate authority and changing pseudonyms (such as [1-3, 9]), therefore smart vehicles equipped with TPDs and vehicles using our solution can coexist and make use of the service provided by the providers. This ensures smooth transition and unlimited overlapping of both solutions.

The certificates can be used for other cryptographic primitives, such as session keys between users, group keys within area for broadcast/multicast of a particular service etc. The solution can guard against Sybil attacks, since one payment processing device will be issued one certificate, if more than one payment device is used then it is possible, but the attacker has to pay for the Sybil node also.

4.4 General Purpose Security Architecture

The security architecture is based on revised Blind scheme. The architecture satisfies required security attributes by using carefully-designed pseudonyms. The pseudonyms are refreshed by vehicles via Roadside Units (RSUs) using revised Blind signature scheme. To refresh pseudonyms, a vehicle uses its previous valid certificate to authenticate its blinded pseudonym-signature-request message to a passing-by RSU. The RSU generates/stores a tag/link based on its received blinded pseudonym-signature-request message and the certificate that was used to authenticate the message.

The tag/link helps to ensure non-repudiation and certificate revocation. It does not require multiple sessions or multiple RSUs to generate this tag/link. The original Blind signature scheme has been modified by enforcing a condition on the blinding factor; this also helps to guard against other attacks towards the original Blind signature scheme (discussed later). It does not generate pseudonyms, with overlapping validity, in bulk which must be guarded against malicious use by user/attacker (e.g., by storing these in a TPD). The non-overlapping pseudonyms or other long term certificates (that may exist at any time) can be securely stored without need of TPD by employing methods that are currently being used in securing certificates in personal computers/servers. The architecture satisfies all security attributes without requiring expensive TPDs or complex multi-step transactions with multiple certificate servers. The architecture does not require users to trust a party with their private/secret keys and thus will have more user acceptance. Further, non-repudiation/revocation requires cooperation between multiple entities thus ensuring privacy without a single point of failure.

4.4.1 Basic Blind Signature - Introduction

Blind signature scheme was first introduced by Chaum [7]. It makes use of multiplicative property of RSA (discussed below). Blind signature scheme based on elliptic curve cryptography can be used interchangeably; we will restrict ourselves to RSA based scheme only.

Entity ***A*** wants to get message *m* blindly signed by entity ***B***; *m* could be hash of some message *M*. Note that the entity ***A*** may need to prove to entity ***B*** that it is entitled to receive blind signatures. The authentication could be done using some token or signatures on message *m*. The details of

authentication are omitted, since it is not essential to the basic concept of Blind signatures. The Blind signature scheme is shown in Figure 4.7.

Given m, s and public parameters, the signatures are valid if $y = m$; where $y = (s)^e = (m^d)^e = m \bmod n$. The Blind signature scheme can be used to certify pseudonyms; but it raises many security issues, such as:

- There is no way to ensure non-repudiation and certificate revocation, since newly signed pseudonyms cannot be linked to authenticator/node (i.e., given $<m, s>$ it is not possible to construct m' or a link to authenticator of m').
- The signed pseudonyms may be used to launch Sybil attacks, and we cannot deal with it since there is no way to link pseudonyms with each other or with the true identity of the node.
- A node with valid authenticator may share its pseudonyms with another node that does not have a valid authenticator and who is unable to get pseudonyms.

A:	1.	Generate random number r: gcd $(r, n)=1$
	2.	Compute blinding factor b_f: $b_f = r^e$
	3.	Blind message m to m' : $m' = b_f m = (r^e m) \bmod n$
$A \rightarrow B$:	4.	m'
B:	5.	Sign message m' using private key d: $x = (m')^d \bmod n$
$B \rightarrow A$:	6.	x
A:	7.	Recover message signature: $s = m^d = r^{-1}(x) \bmod n$
	8.	$r^{-1}(x) \bmod n = r^{-1}(m')^d \bmod n = r^{-1}(r^e m)^d \bmod n$ $= r^{-1} r\, m^d \bmod n = m^d \bmod n$

Figure 4.7: Basic Blind signature scheme (public key parameters: n, e = public key of B and d= secret key of B).

4.4.2 Proposed Architecture

The architecture uses a certificate chain consisting of long-term and short-term certificates. Initially, a long-term certificate is used to get the initial short-term certificate and later a new short-term certificate can be obtained based on the previous short-term certificate, thus making a certificate chain (details discussed later in this section). We have revised the Blind signature scheme to meet our requirements of non-repudiation and revocation.

4.4.2.1 Notations and Function Definitions

We define several notations and functions which we will use in formal description of our architecture. A certificate or pseudonym $Cert_x$ is essentially defined by its associated identification ID_x and key pair (P_x, S_x); public key P_x forms part of certificate and secret key S_x is known only to the holder of $Cert_x$. *Sig-*$Cert_x$ *(M)=N,* is a signature function on message *M* using key S_x or certificate $Cert_x$. The signature *N* is computed by first creating a message digest (M_h = *Hash(M)*) using some well known hashing function (such as SHA1) and then encrypting the digest using key S_x. *VerSig-*$Cert_x$ *(M, N)*, is a signature verification function with two inputs: the message *M* and the signature *N*. It verifies the signature by computing the message digest of message *M* and comparing it with received signature *N* (after decrypting it with the corresponding public key P_x). Note that knowledge of $Cert_x$ is needed for function *VerSig-*$Cert_x$ *(M, N)*. $Cert_x$ should either be publicly available or attached along with the *Sig-*$Cert_x$ *(M)*. In the rest of the paper it is assumed that $Cert_x$ is either publicly available or attached along with the *Sig-*$Cert_x$ *(M)*, and will not be explicitly mentioned.

4.4.2.2 Proposed Revised Blind Signature Scheme

In order to address the security issues of the original Blind signature scheme and to satisfy our security objectives, we revised Blind signature scheme (Figure 4.8). Our proposed scheme achieves *one-way-link-ability,* i.e., given a blinded pseudonym (*m'*) the signer cannot find the un-blinded pseudonym (*m*), but given a certificate (<un-blinded pseudonym -*m,* signatures -*s*>) the signer can construct the associated blinded pseudonym (*m'*) and find a link to its authenticator (authenticator of *m'*). *One-way-link-ability* ensures privacy since the signer, at the time of signing signatures, cannot determine the pseudo-credentials. Whereas for revocation/non-repudiation (given the pseudonym), it is possible to construct the chain/link leading to the node's true identity.

Suppose that vehicle *V* has a current certificate $Cert_{i-1}$ which is valid for time period T_{i-1} (time period defines a start and an end time) and now needs to get a new certificate $Cert_i$ valid for time period T_i from a nearby RSU *R* (Fig. 2). *V* generates $Cert_i$ (step 1), blinds the certificate using public credentials of RSU *R* (steps 2, 3), authenticates the blind-signature-request-message with $Cert_{i-1}$ and sends the request to the RSU (step 4). RSU *R* verifies validity of $Cert_{i-1}$ (step 5), verifies signatures on request (step 6), generates/stores the tag/link (steps 7, 8) and sends signed message to *V* (step 9). *V* un-blinds the message to get the signatures on pseudonyms (step 10) and then uses the pseudonym as required (step 11), but makes sure to not use $Cert_i$ with *R*.

The solution ensures *one-way-link-ability* to achieve non-repudiation/revocation: $Cert_i$ cannot be generated from m_i', but m_i' can be generated from $Cert_i$ and m_i' can be linked to $Cert_{i-1}$. Note that revocation/non-repudiation cannot be accomplished by a single signing RSU (or a few RSUs); it

requires cooperation between all involved RSUs to reconstruct the chain/link iteratively. The utility of *one-way-link-ability* rests on the assumption that a node should not declare (use) the un-blinded pseudonym to (with) the RSU who issued signatures on its blinded version.

V:	1.	Generate $Cert_i = < ID_i , P_i >$ and S_i ; $ID_i = b_f = r^e$; where ID_i is pseudo ID, P_i and S_i are public and secret keys of *V*.
	2.	Compute $m_i = Hash\ (Cert_i)$
	3.	Compute $m_i' = b_f\, m_i = ID_i m_i = (r^e\ m_i) \bmod n$
$V \rightarrow R$:	4.	m_i', T_i , $Sig\text{-}Cert_{i-1}\ (m_i')$, $Cert_{i-1}$
R:	5.	Verify $Cert_{i-1}$ for time-period validity and revocation. (details in certificate revocation)
	6.	$VerSig\text{-}Cert_{i-1}\ (m_i', Sig\text{-}Cert_{i-1}\ (m_i'))$
	7.	Compute $x = (m_i')^d \bmod n$
	8.	Store link $<m_i', T_i , Cert_{i-1}>$ *or* alternatively $< m_i', T_i , m_{i-1}' >$
$R \rightarrow V$:	9.	$x, T_i , Sig\text{-}Cert_R\ (T_i)$
V:	10.	Recover certificate signature $Sig\text{-}Cert_R\ (Cert_i)$: $Sig\text{-}Cert_R\ (Cert_i) = s = (m_i)^d = r^{-1}(x) \bmod n$
	11.	Use $<Cert_i , T_i , Sig\text{-}Cert_R(Cert_i), Sig\text{-}Cert_R(t_n), Cert_R>$ as new credentials

Figure 4.8: Proposed revised Blind signature scheme – initial version (public key parameters: n, e = public key of *R*, d = secret key of *R*).

The solution shares a limitation with Blind signature scheme, i.e., the signer cannot ensure that the blinded message (certificate) is well-formed (constructed as per agreed protocol/scheme). Specifically the RSU cannot ensure $b_f = ID_i$. One commonly used solution is to use cut-and-choose method [4, 6]. Here the user sends multiple certificates to the signer (e.g., user sends two blinded messages, if she wants to get signatures on one); the signer can then choose which half to sign and the user un-blinds the other half for the signer to check if these were well-formed or not. This

reduces the success probability of attacks by malicious users but at the same time adds considerable overhead, which is not affordable in the face of intermittent connectivity in VANET environment. In order to address this vulnerability, we have further refined the Blind signature scheme. The modifications are given in Figure 4.9 (only shows the several revised steps in the initial proposed approach given in Figure 4.8).

$Sig\text{-}Cert_R\ (m_i'_H\ ||\ T_i)$ is the modified Blind signature, which serves three purposes: attaching a certificate-valid-time-period condition to the signature, adding link-ability to the certificate for later non-repudiation/revocation purpose, and guarding against malicious use of the signature (malforming the blind message, changing ID to make certificate un-linkable, sharing the signed certificate, etc). [1]

R:	7.	Compute $m_i'_H = Hash\ (m_i')$
	8.	Store link <mi' H, Ti , Certi-1> or alternatively < mi'H , Ti , mi-1'H >
$R \rightarrow V$:	9.	$m_i'_H$, T_i , $Sig\text{-}Cert_R\ (m_i'_H\ ||\ T_i)$; $x \| y$ is concatenation of x and y.
V:	10.	Use <$Cert_i$, T_i , $Sig\text{-}Cert_R(m_i'_H\ ||\ T_i)$, $Cert_R$> as new credentials

Figure 4.9: Proposed revised Blind signature scheme – final version.

[1] To guard against blind decryption or blind signatures on some other message besides certificates, certification servers will have different certificates for signing and encrypting other messages.

4.4.3 System Setup

Three types of certificates have been defined: permanent certificates, long-term/daily certificates, and short-term/temporary certificates (i.e., pseudonyms) (Figure 4.10). Each vehicle will have a permanent certificate that is registered with a Central Certification authority (CCA) similar to vehicle registration authority. The CCA can be state or country based and its operational area is divided into regions, with each region having a Regional Certification Authority (RCA). A vehicle on entering a region registers itself with the RCA; RCA in turn updates the vehicle's current region information on CCA (the update only takes place when a vehicle moves from one region to another). Either RCA or CCA can confirm that the permanent certificate of vehicle has not been previously revoked. This helps to target the revocation to concerned regions only, and hence, simplifies revocation and reduces certificate revocation list (CRL) size. The size of a region depends on the desired privacy granularity.

A vehicle gets one long-term certificate per day from RCA using proposed Blind signature scheme. RCA stores the relevant link. One long-term certificate per day reduces the chain size which makes revocation simple. A vehicle uses this long-term certificate to get its first short-term certificate (of the day) from an RSU, using the modified Blind signature scheme proposed in this paper. The RSU stores the relevant link/tag in its database and informs RCA (via a confirmation message) that a short-term certificate has been issued based on a particular long-term certificate. The RCA modifies the freshness/used bit associated with the record. If later the RCA receives another certificate-issue-confirmation-message for the same long-term certificate, it marks the vehicle as malicious and takes

appropriate measures such as certificate revocation. RSUs can use $m_i'_H$ instead of un-blinded long-term certificate in confirmation message to further ensure privacy.

For each subsequent certificate, the vehicle uses its previous/last short-term certificate to authenticate its current request. The issuing/signing RSU in this case sends a confirmation message to the RSU who issued/signed the previous short-term certificate. The RSU who issued/signed the previous short-term certificate then modifies the freshness/used flag associated with the record. This ensures that more than one certificate are not issued based on one particular short-term certificate. The time period of the new certificate will be non-overlapping and later than the validation period of the previous certificate.

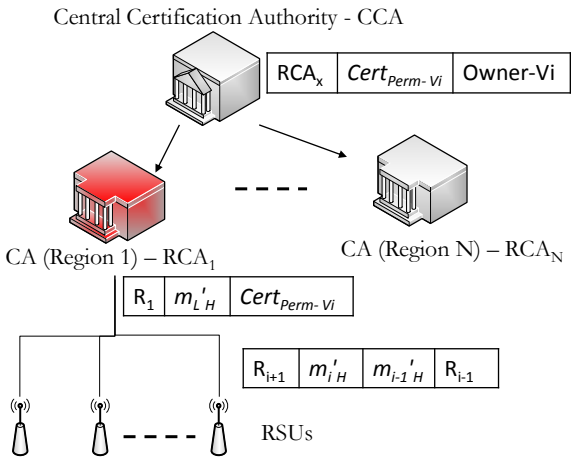


Figure 4.10: Certificate Architecture: CCA maintains current RCA and permanent certificate to owner link, RCA maintains permanent certificate to blinded-long-term certificate link and RSU that reported usage of long-term certificate, each RSU maintains authenticating-certificate (and its issuer) to blinded-short-term certificate link and the RSU that reported usage of issued short-term certificate.

The issuer/signer (RCA/RSU) of authenticating certificate ($Cert_{i-1}$) while modifying the freshness/used bit may also record the RSU which has sent the confirmation message. This will simplify revocation process (discussed later in Section 4.4.4) but will also raise limited privacy issues since the RSU knows the link to the next RSU as well as the previous RSU. Note that even with this knowledge a single RSU cannot compromise the privacy of a vehicle; it still needs cooperation from other RSUs, though in this case it knows the RSUs with which it should cooperate.

We require RSUs to send certificate-issue-confirmation-message to issuer/signer (RCA/RSU) of authenticating certificate ($Cert_{i-1}$). This ensures that no more than one pseudonym with same/overlapping validity will be issued. For this goal, the RSU sends certificate-issue-confirmation-message to the issuer/signer (RCA/RSU) of authenticating certificate ($Cert_{i-1}$) and waits for a timeout before signing new pseudonym. Issuer/signer (RCA/RSU) of authenticating certificate ($Cert_{i-1}$) responds within the timeout period only if malicious activity is detected. This ensures desired security with minimum overhead.

4.4.4 Security Attributes

Confidentiality, integrity, and authentication can be achieved by using short-term certificates for signatures and/or encryption. Rest of the security attributes are discussed below:

4.4.4.1 Privacy

The solution ensures privacy since the RCA/RSUs do not know the ID and public keys of a vehicle at the time of signing the blinded certificate. Further, since a vehicle gets a short-term certificate from one RSU and uses it later with another RSU, a single RSU cannot link different short-term certificates of a particular vehicle. Tracing is possible but quite difficult for attackers, which can be achieved only when all the RSUs that issued certificates to a particular vehicle cooperate with each other. Even if attackers can trace a vehicle in this way, the true ID of a vehicle cannot be determined without the help from RCA. Also, RCA by itself cannot compromise the privacy of the vehicle. This property improves users' confidence since even the government authority itself cannot compromise user privacy---government authority must get cooperation from commercial operators who operate the RSUs in order to trace a vehicle and its user.

4.4.4.2 Non-repudiation (Liability)

If a malicious message, signed by a particular certificate, has been identified then the privacy of signer can be revoked with the cooperation between RCA and RSUs. The signed message will contain the information < $Cert_i$, T_i , $Sig\text{-}Cert_{Rn}$ ($m_i'_H$ || T_i), $Cert_{Rn}$>. It is assumed that the certificate and signatures on the malicious message are valid, since if the certificate is not valid then the message will be discarded and there will be no need of revocation. The revocation is performed backwards iteratively as following (refer to Figure 4.11):

- RSU – R_n will generate $m_n'_H$, locate the record and find corresponding $Cert_{n-1}$.
- It will then forward the revocation request to RSU-R_{n-1} which issued $Cert_{n-1}$.

- The chain will be followed till first RSU and then RCA which will reveal the true ID of malicious vehicle (based on long-term certificate).

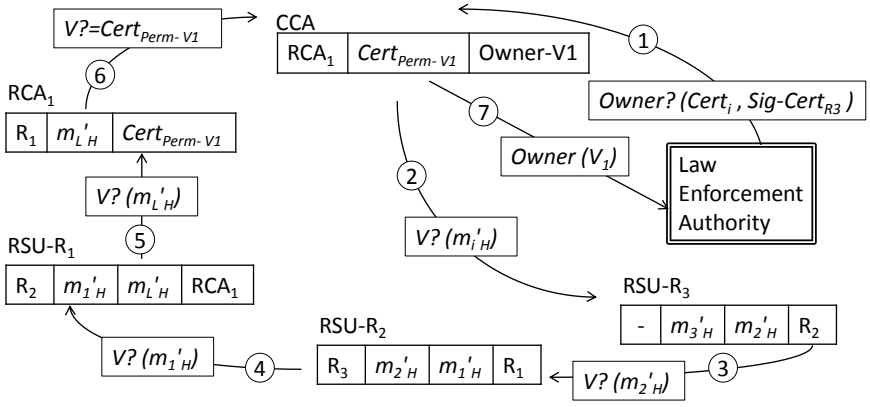


Figure 4.11: Non-repudiation procedure: (1) Law enforcement forwards the short-term certificate under investigation to CCA, (2) CCA forwards the blinded-short-term certificate to the concerned RSU, (3,4) RSUs iteratively forward the blinded-authenticating-short-term certificate to its issuing RSU, (5) RSU forwards the blinded-long-term certificate to RCA, (6) RCA forwards corresponding permanent certificate to CCA, (7) CCA provides the ownership information to requesting Law enforcement authority.

4.4.4.3 Certificate Revocation

There may be a situation when the certificates of a particular vehicle are to be revoked. It is important to note that the majority of vehicles will be honest and certificate revocation will not be routine so the proposed protocol has been designed to minimize overheads in normal situations. The certificate revocation decision may be made at CCA based on either request of user (for stolen credentials) or law enforcement (for malicious use). The detail of decision methodology is out of the scope of this paper and will not be discussed. The CCA will inform the RCA of the region where the vehicle last registered. Revocation is processed iteratively along the certificate chain in forward direction as following (Figure 4.12):

- RCA will check to see if the vehicle has already used its long-term certificate (to get short-term certificate from some RSU) or not. If the vehicle did not get any short-term certificate then RCA will revoke long-term certificate by broadcasting revocation command containing the hash of the blinded long-term certificate ($m_i'_H$). RSUs will not issue first short-term certificate based on this long-term certificate. The certificate revocation command will expire after the validity of long-term certificate.
- If the vehicle has used its long-term certificate to get the short-term certificate then the RCA will broadcast revocation command to all RSUs containing hash of corresponding blinded short-term certificate ($m_i'_H$). Note if the RCA maintains the ID of RSU that issued the first short-term certificate (based on confirmation message sent by the RSU) then the revocation command is only needed to sent to the single RSU that issued the first short-term certificate.
- The RSU that issued the first short-term certificate will find the link and broadcast the revocation command containing hash of corresponding blinded short-term certificate ($m_i'_H$). RSU may also acknowledge to the RCA. The broadcast may be limited to a few hops since it is likely that vehicle would have got the next short-term certificate from some RSU in geographical proximity of the first RSU. The broadcast range may be expanded if no RSU acknowledges. Similarly if RSUs maintain the ID of the next RSU in certificate chain then revocation message may be sent directly to the concerned RSU.
- The revocation broadcast ends at the last RSU that issued the short-term certificate, the RSU then broadcasts revocation message containing hash of current blinded short-term certificate ($m_i'_H$). Other RSUs will not issue any new certificate based on this short-term certificate and will also not trust any message signed by this certificate. RSUs may also

broadcast hash of current blinded short-term certificate ($m_i'_H$) to vehicles in the limited region (the limit can be defined) within the validity time period of certificate.

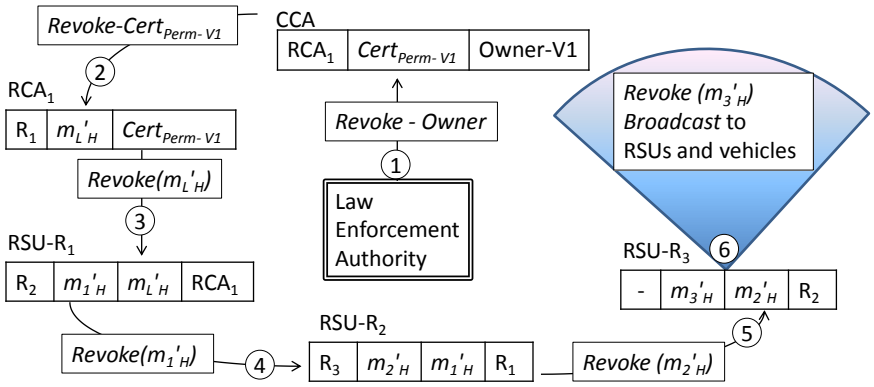


Figure 4.12: Revocation Procedure: (1) Law enforcement authority forwards ownership information, (2) CCA forwards the permanent certificate to concerned RCA for revocation, (3) RCA forwards the blinded-long-term certificate to concerned RSU, (4,5) RSUs iteratively forward the blinded-short-term certificate to next RSU that reported its usage as authenticating certificate (6) Last RSU broadcasts the blinded-short-term certificate to all RSUs and nearby vehicles.

4.5 Conclusion

Two different security architectures have been presented to address the security challenges during initial deployment stage of VANET. The service oriented security architecture can be incrementally deployed. Users are issued with temporary certificates which can only be used within a specific geographic area and within a particular time period. This property also simplifies the certificate revocation procedure. A framework has been presented which can be used to provide various services to VANET by providers without investing much in infrastructure. The solution is intended to stimulate people's interest in VANET and build user/provider confidence. The general purpose security architecture is based on revised Blind signature scheme. The Blind signature scheme has

been revised to ensure provision of all the security attributes. The solution does not require tamper-proof-devices or multiple interactive transactions. Non-repudiation/revocation requires cooperation between multiple entities thus ensuring privacy without a single point of failure.

4.6 References

[1] P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya, Z. Ma, F. Kargl, A. Kung, and J. P. Hubaux, "Secure vehicular communication systems: design and architecture", In IEEE Wireless Communication Magazine, November 2008.

[2] F. Armknecht, A. Festag, D. Westhoff, and K. Zang, "Cross-layer privacy enhancement and non-repudiation in vehicular communication", In Proceedings of the 4th Workshop on Mobile Ad-Hoc Networks (WMAN'07), March 2007.

[3] C. I. Fan, R. H. Hsu, and C. H. Tseng, "Pairing-based message authentication scheme with privacy protection in vehicular ad hoc network", In Proceedings of the International Conference on Mobile Technology, Applications and Systems, September 2008.

[4] S.U. Rahman and U. Hengartner, "Secure crash reporting in vehicular ad hoc networks", In SecureComm'07.

[5] L. Fischer, A. Aijaz, C. Eckert, and D. Vogt. "Secure Revocable Anonymous Authenticated Inter-Vehicle Communication", ESCAR'06.

[6] F. Schaub, F. Kargl, Z. Ma, and M. Weber, "V-tokens for Conditional Pseudonymity in VANETs", ", In IEEE WCNC'10.

[7] D.Chaum, "Blind signatures for untraceable payments", In Advances in Cryptography, CRYPTO 82, PlenumPress. 1983: pp.199-20.

[8] X. Lin, X. Sun, P. H. Ho, and X. Shen, "GSIS: a secure and privacy preserving protocol for vehicular communications", IEEE Transaction on Vehicular Technology, Vol.56, No.6, pp.3442-3456, 2007.

[9] "IEEE P1609.2 trial-use Standard for Wireless Access in Vehicular Environments: Security Services for Applications and Management Messages," July 2006.

[10] M. Raya, P. Papadimitratos, and J.-P. Hubaux, "Securing Vehicular Communications," In IEEE Wireless Communications Magazine, pp 8-15, October 2006.

[11] P. Papadimitratos, L. Buttyan, J-P. Hubaux, F. Kargl, A. Kung, and M. Raya, "Architecture for Secure and Private Vehicular Communications," In Proceedings of the 7th International Conference on ITS Telecommunication, June 2007.

[12] G. D. Crescenzo, T. Zhang , and S. Pietrowicz, "Anonymity notions for public-key infrastructures in mobile vehicular networks", In Proceedings of the 4th IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS'07), October 2007.

[13] J. Choi and S. Jung, "A security framework with strong non-repudiation and privacy in VANETs", In Proceedings of the 6th Annual IEEE Consumer Communications & Networking Conference IEEE CCNC 2009, January 2009.

[14] A. Stampoulis and Z. Chai. Survey of security in vehicular networks, project CPSC 534, http://zoo.cs.yale.edu/ ~ams257/projects/wireless-survey.pdf, 2007, last access: May 14th , 2009.

[15] P. Papadimitratos, G. Mezzour, and J. P. Hubaux, "Certificate revocation list distribution in vehicular communication systems", In Proceedings of the fifth ACM international workshop on VehiculAr Inter-NETworking VANET'08, September 2008.

[16] B. Parno and A. Perrig, "Challenges in Securing Vehicular Networks", In Proceedings of the Workshop on Hot Topics in Networks (HotNets-IV), November 2005.

[17] X. Lin, R. Lu, C. Zhang, H. Zhu, P. H. Ho, and X. Shen, "Security in vehicular ad hoc networks", In IEEE Communication Magzine, pp 88-95, April 2008.

[18] L. Buttyan, T. Holczer, and I. Vajda, "On the effectiveness of changing pseudonyms to provide location privacy in VANETs", In Proceedings of the 3rd European Workshop on Security and Privacy in Ad hoc and Sensor Networks, July 2007.

[19] A. R. Beresford and F. Stajano, "Mix Zones: User privacy in location-aware services", In Proceedings of the 1st IEEE International Workshop on Pervasive Computing and Communication Security (PerSec), March 2004.

[20] S. Eichler, "Strategies for pseudonym changes in vehicular ad hoc networks depending on node mobility", In Proceedings of the 2007 IEEE Intelligent Vehicles Symposium, June 2007.

[21] M. Gruteser and D. Grunwald, "Enhancing location privacy in wireless LAN through disposable interface identifiers: A quantitative analysis", Mobile Networks and Applications, v.10 n.3, p.315-325, June 2005.

[22] E. Fonseca, A. Festag, R. Baldessari and R. Aguiar, "Support of anonymity in VANETs – Putting pseudonymity into practice", In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), March 2007.

[23] J. Freudiger, M. Raya, M. Felegyhazi, P. Papadimitratos, and J.-PHubaux, "Mix-zones for location privacy in vehicular networks", In Proceedings of the 1st International Workshop on Wireless Networking for Intelligent Transportation Systems (Win-ITS), August 2007.

CHAPTER 5 OPTIMAL PLACEMENT OF RSUs

Vehicular ad hoc network (VANET) has numerous applications and most of these applications collect/disseminate information from/to vehicles. The collection/dissemination of this information from/to vehicles takes place via roadside units (RSUs). The effectiveness of this information flow depends on connection time/bandwidth. Connection time, between vehicles and RSUs, can easily be improved by pervasive deployment of RSUs. However, this is an expensive solution and will especially not be feasible during the initial stages of VANET when market penetration will be very low. The focus is on applications that collect information, such as information about road conditions, traffic conditions or traffic accident, from vehicles to nearby RSU system. The optimization goal is to minimize the average reporting time for all possible information reports in a local region; reporting time is defined as the time duration from occurrence of an event till it is reported by a vehicle to an RSU. The proposed optimization scheme can easily be extended to applications that disseminate information, here the optimization goal can be the area covered within some time bounds.

The chapter is organized in four sections. Section 5.1 discusses the related research work. Section 5.2 presents optimal placement of RSUs along highways. Section 5.3 discussed the optimal placement of RSUs in Urban areas. Section 5.4 concludes the chapter.

5.1 Related Work

Earlier works in optimal placement in VANET include [3-11]. Lee et al. [3] seek optimal placement of RSUs to improve connectivity. Each intersection is considered as a potential RSU location. These potential locations are then ordered based on number of vehicle-reports received within communication range of each RSU. The vehicle-reports are based on per minute locations reported by taxis to telematics system. RSUs' locations are then selected from this ordered list. The placement scheme only considers taxi location reports and does not consider speed or density of all vehicles. It, therefore, may not be able to achieve optimal connectivity for all vehicles.

Li et al. [4] consider the optimal placement of gateways, which connect RSUs (access points - AP) to the Internet, while minimizing the average number of hops from APs to gateways. They consider pervasive APs such that every vehicle is connected to an AP. They do not consider vehicle speed, density or movement patterns.

Zhao et al. [5] optimize placement of Thowboxes, standalone units that act as relays, to improve contact and data-rate/throughput within context of a delay tolerant network. They aim at improving V2V communication and not the V2I communication.

Lochert et al. [6] use genetic algorithm for optimal placement of RSUs for a VANET traffic information system. They use a hierarchical aggregation scheme to share traffic information among vehicles and the optimal placement is aimed at minimizing the travel time based on this information

sharing. The optimal placement is to minimize travel for some fixed landmarks and may not be useful for travel between any two points in an area.

Sun et al. [7] optimize the location of RSUs such that vehicle can reach an RSU within some timing constraint, given by sum of driving time and an overhead time (for adjusting the route), to update short term certificates. The optimization scheme may require vehicles to change their route which may have effects on local traffic condition. We do not have any route changing condition; we optimally place the RSUs considering the vehicles current routes only.

Fiore et al. [8] optimally place RSUs (Access Points - AP) in an urban environment to improve cooperative download of data among vehicles. They aim at placing the APs at point where maximum vehicles cross each other, this helps in relaying the data from AP to a downloading vehicle via other vehicles. Trullols et al. [9] optimally deploy RSUs (Dissemination Points – DPs) in an urban area to maximize the number of vehicles that contact the DPs. They also consider a second case where, in addition to the number of vehicles that contact DPs, the contact times of vehicles are also taken into consideration. Malandrino et al. [10] optimally deploy the RSUs (APs) to maximize the system throughput. They consider both the V2I (or I2V) and V2V communications for optimal placement of APs. Vehicle trajectory information (time and location) forms basis of this optimization which may not be available in many cases. Zheng et al. [11] optimally deploy APs to improve contact opportunity; defined in terms of time for which a user remains in contact with an AP. A moving user may connect with different APs during different times whereas; we only consider the time it takes for the vehicle to report an incident to first RSU it encounters. These optimizations aims at transfer of data from RSUs to vehicles whereas, our optimization aims at

transfer of data from vehicles to RSUs with an area coverage constraint. Also, we do not consider V2V communication in our optimization problem.

The work is also related to the problem of facility location, where one or more facilities are optimally located in a region to reduce the overall costs (to consumer and facility) [12, 13]. The work does not aim at minimizing the overall costs (reporting time of events) rather it aims at minimizing the average reporting time on each path/route basis; this need awareness to road topology. Further, It also incorporates vehicle speed, vehicle density, probability of a vehicle to follow a particular route and event distribution.

5.2 Optimal Placement of RSUs along Highways

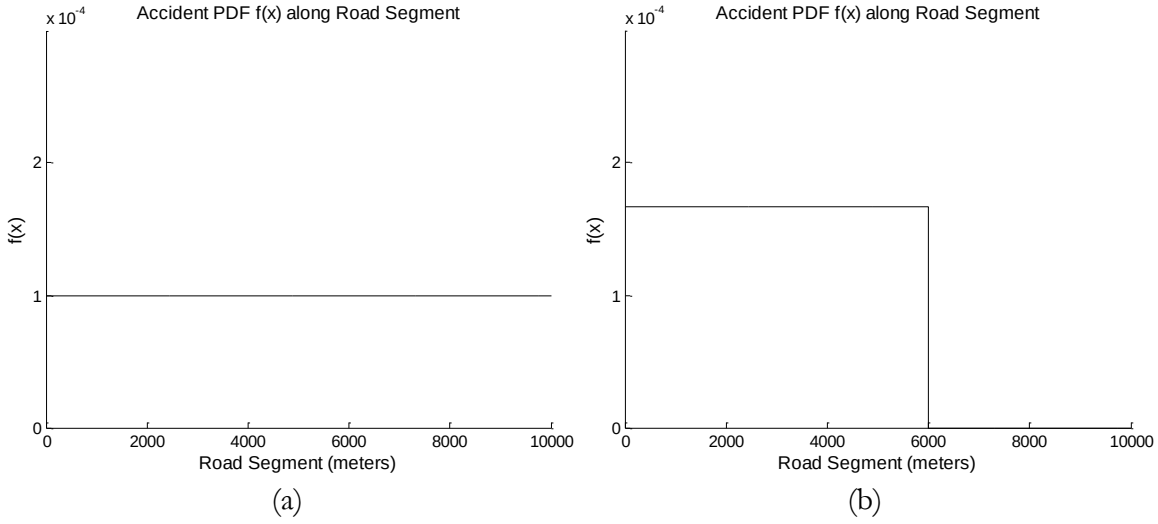
Given a limited number of RSUs, the section addresses the issue of optimal placement of these RSUs along highways with the goal of minimizing the average time taken for a vehicle to report an event of interest to a nearby RSU. One obvious solution is to uniformly distribute the available RSUs along the highway. This solution may be effective where the need of information collection/dissemination is uniform along the whole road range of a highway, which may not always be the case. For example, if we are interested in collection of information about road conditions such as fog or ice, then some areas will always be more likely to have a fog or an ice condition than the other areas. Finding the optimal solution via exhaustively checking all possible placement strategies will become infeasible with the increase of the number of RSUs, e.g., on a 100Km highway we can have approximately 200 candidate locations for RSUs (if RSU communication range is 250m

and an RSU can only deployed after every 2x250m), and if we need to place 20 RSUs among these locations then there will be 1.61 x 10^{27} different placement strategies.

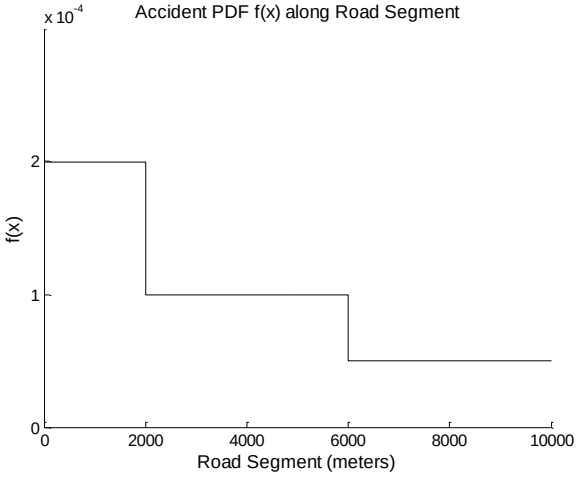
5.2.1 System Model

The scope of this paper is restricted to optimal placement of RSUs along one single highway. Let L be the length of a highway and R be the communication range of an RSU/vehicle. If RSUs can only be deployed after every 2xR distance then there will be $N = \lfloor L/(2R) \rfloor$ candidate locations. If M is the number of available RSUs then we aim at placing these M RSUs among N locations such that the average reporting time $T(X)$ is minimized. $X=\{x_1, x_2,.., x_M\}$ and x_i is the location of RSU_i.

The density and speed of vehicles along the highway is denoted by $d(x)$ and $s(x)$ respectively, where $0 \leq x \leq L$. For simplicity we consider a constant density D and constant speed S. If vehicles entering the highway follow Poisson distribution then there will be $\lambda = SD$ vehicles entering the highway per unit time. If y is the location of an incident/event that needs to be reported to RSU systems, a vehicle will reach the point of incident with an exponentially distributed time (mean value is $1/\lambda$). Let $f(x)$ define the distribution function of incidents/events along the highway. In order to simplify evaluation, we have considered $f(x)$ such that the locations of optimal RSUs can be derived intuitively (the three scenarios shown in Figure 5.1).



(a) (b)



(c)

Figure 5.1: Incident/event distribution functions. (a) Flat (b) Step (c) Stair

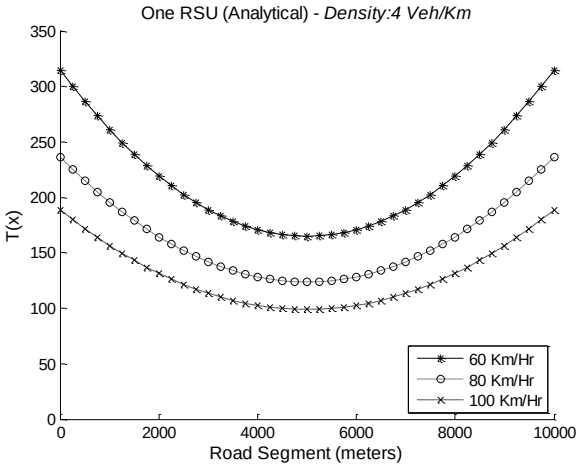
5.2.2 Simple (Analytical) Optimization

If an RSU is located at position x, then for an incident/event that happened at place y, the reporting time $t(x|y)$ is the summation of the time for a vehicle to arrive at y (denoted as t_y) and the time for the vehicle to reach x from y (denoted as $t_{y\to x}$), see Equation 5.1. If density and speed for vehicles in both directions are same then the average reporting time will be given by Equation 5.2. The

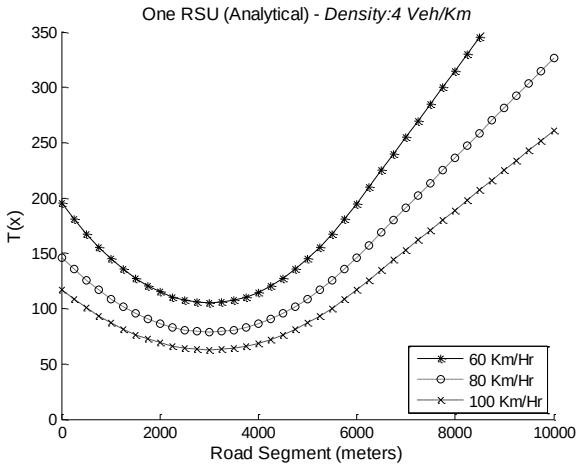
formulas for other cases such as different speed/density on opposite directions and variable speed/density have been omitted due to space limitations. For *M=1* (single RSU placement problem), *T(x)* for all possible locations of the single RSU is shown in Figure 5.2. The optimal position of the RSU should have the minimum *T(x)*. For *M=2,* optimal positions of the two RSUs are shown in Figure 5.3.

$$t(x|y) = t_y + t_{y\to x} = \frac{1}{SD} + \frac{|x-y|}{s} \tag{5.1}$$

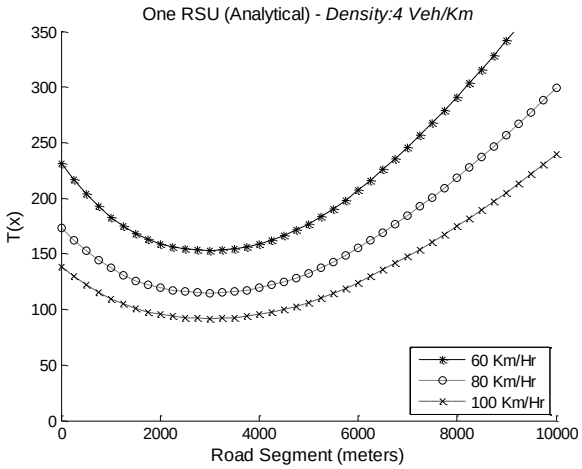
$$T(x) = \int_0^L t(x|y)f(y)dy \tag{5.2}$$



(a)

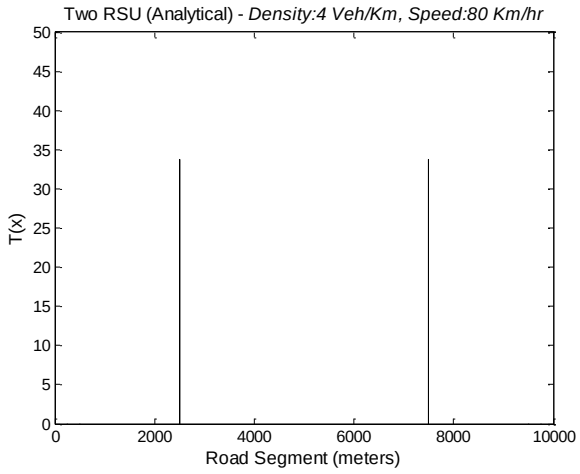


(b)

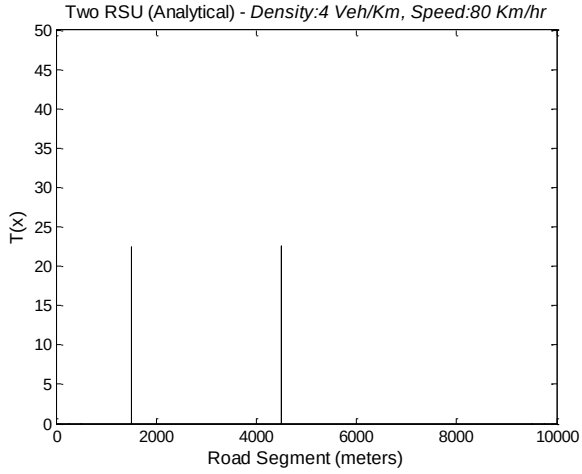


(c)

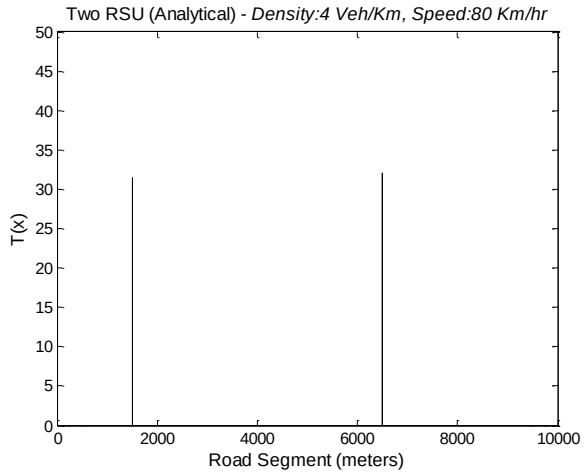
Figure 5.2: T(x) (Simple method) for M=1. (a) Flat (b) Step (c) Stair



(a)



(b)



(c)

Figure 5.3: Optimal RSU positions (Simple method) for M=2. (a) Flat (b) Step (c) Stair

5.2.3 Balloon Optimization

In this optimization method RSUs are considered as balloons, where the balloon boundaries represent the coverage area of an RSU. Let ***a*** and ***b*** be the balloon boundaries of an RSU such that $0 \leq a \leq b \leq L$. The average reporting time for each side is considered independently; $T_a(x)$ and $T_b(x)$ for side bounded by a and b respectively, and their formulas are:

$$T_a(x) = \int_a^x \left(\frac{1}{SD} + \frac{(x-y)}{S}\right) f(y)dy, \quad T_b(x) = \int_x^b \left(\frac{1}{SD} + \frac{(y-x)}{S}\right) f(y)dy \tag{5.3}$$

Initially, RSUs are positioned uniformly along the highway, with $a=b=x$ and $T_a(x)=T_b(x)=0$. In each optimization iteration, the $T(x)$ on both sides of each balloon is incremented by a small value; each balloon is then expanded independently on both sides (i.e., a and b are increased) such that the computed value of $T(x)$ on both sides equals the newly incremented value. Note that the expansion of a balloon on both sides may not be uniform. For example, the side with lower values of $f(x)$ will expand more. The process is repeated till the balloon touches another RSU/balloon or the highway

boundary. The expanded balloons are then repositioned such that each balloon is equidistant from other balloons or highway boundaries (in a similar way like multiple balloons bounce with each other in a constrained space). The process is then repeated all over again. The process continues till there is no more space for expansion of balloons. The positions of RSUs at this point is the optimal solution. Optimal RSU positions for *M=1* and *M=2* are shown in Figure 5.4 and Figure 5.5.

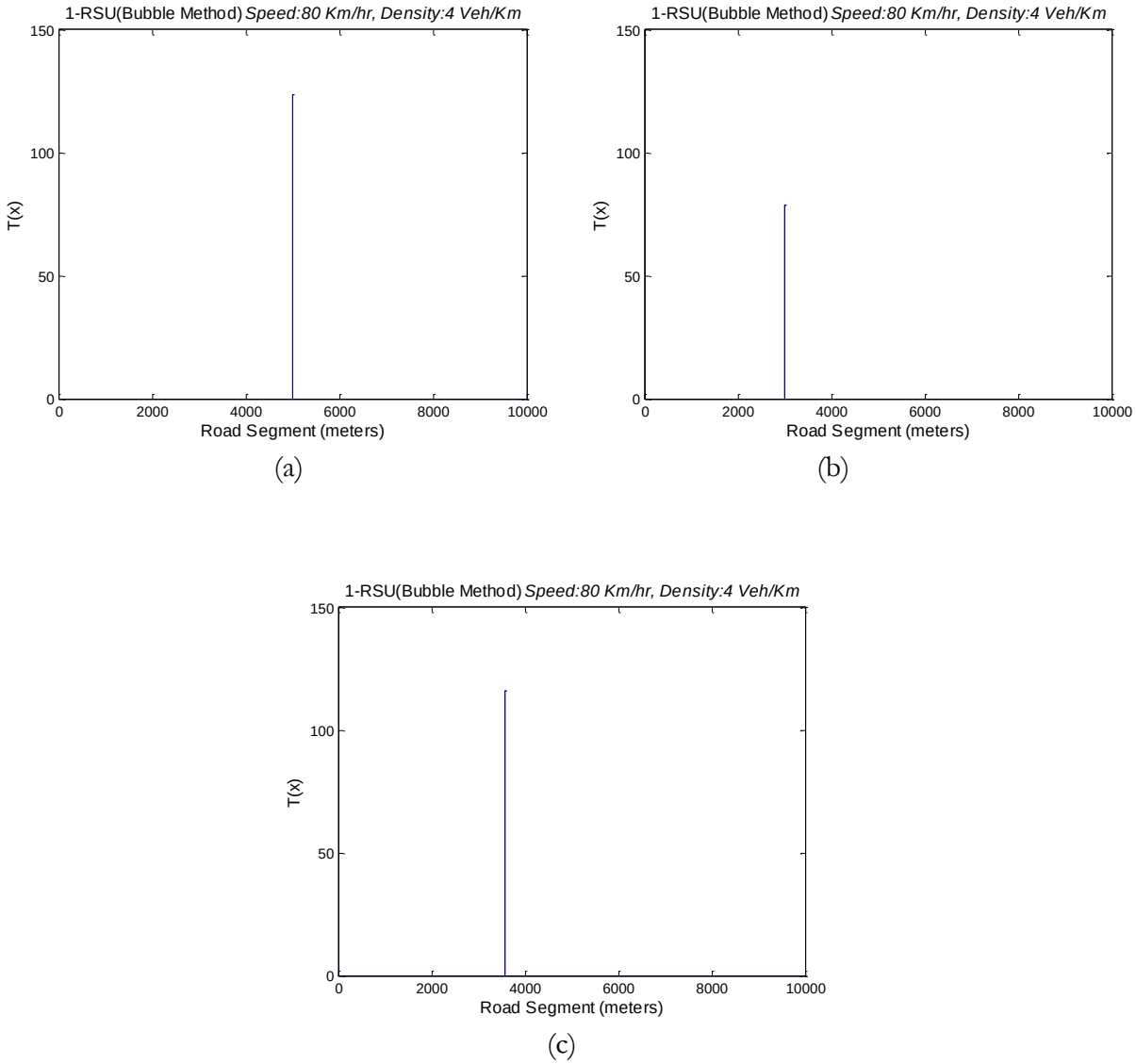


Figure 5.4: Optimal RSU positions (Balloon method) for M=1. (a) Flat (b) Step (c) Stair

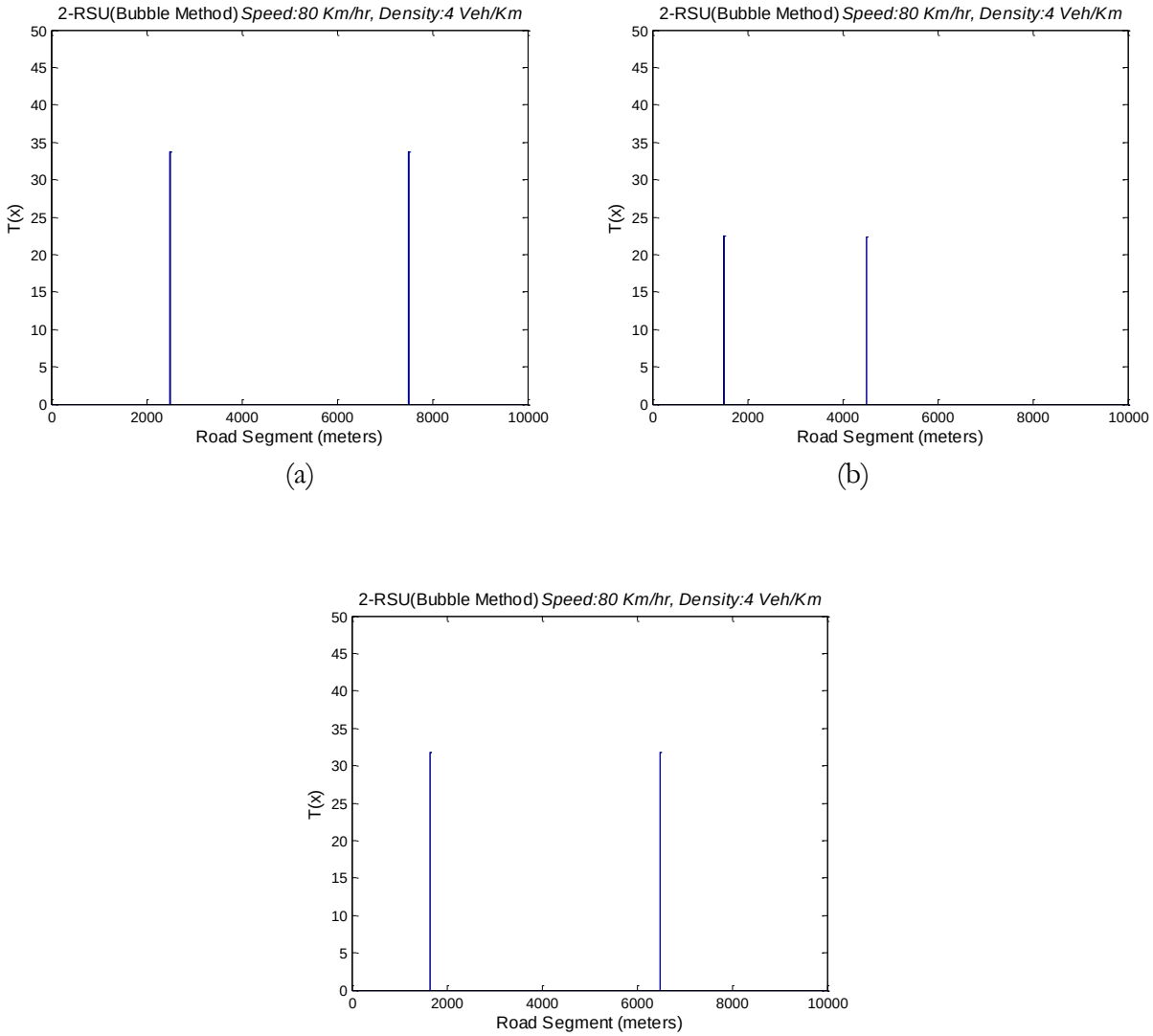


Figure 5.5: Optimal RSU positions (Balloon method) for M=2. (a) Flat (b) Step (c) Stair

5.2.4 Discussion

As shown in Figure 5.1, three different distributions of *f(x)* were especially chosen so that the optimal positions of RSUs are intuitive. Figure 5.2 and 5.4 show optimal positions for *M=1* and Figure 5.3 and 5.5 show optimal positions for *M=2*. The simple (analytical) optimization method finds optimal placement after exhaustively going through all the possible options (see Figure 5.2)

and will thus be uneconomical for long highways or a large number of RSUs. Balloon optimization heuristics on the other hand is much simpler and less complex. The optimal positions from both the optimization methods closely match with each other and are also intuitive.

5.3 Optimal Placement of RSUs in Urban Areas

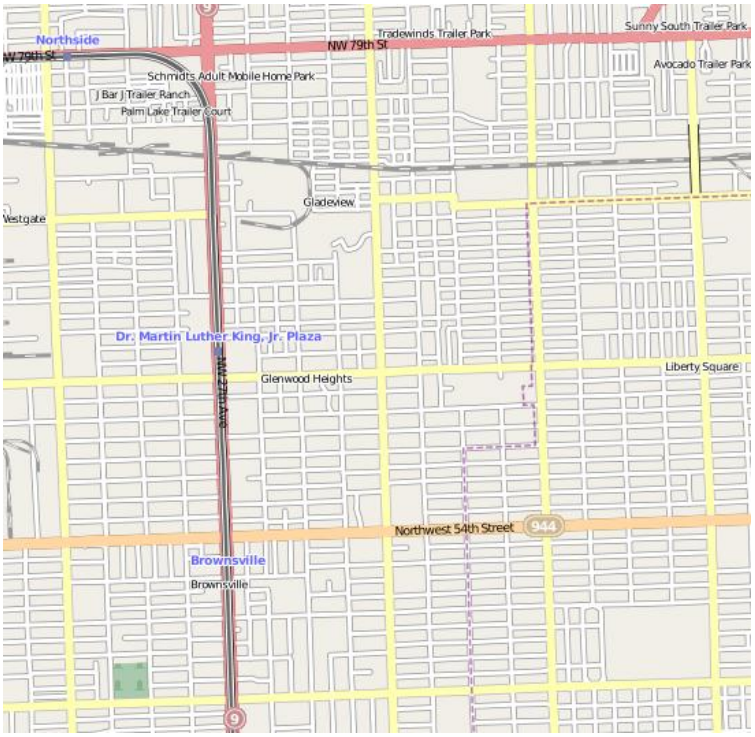
The scope of this part is restricted to urban environment such as the one shown in Figure 5.6. Figure 5.6(a) shows a partial map of Miami, FL, USA. The map shows a grid of major roads (shown in yellow and red color) and a number of smaller/local streets. The major roads are shared by all users/buses for commuting whereas the smaller streets are used only by users who need to visit a particular home or business on that street. The traffic on smaller streets is therefore very small/negligible as compared to that on major roads and we can safely ignore these for our system model. Figure 5.6(a) can be approximated to a grid network of roads as shows in Figure 5.6(b) after removing the local/smaller streets.

5.3.1 Optimization Problem Modeling

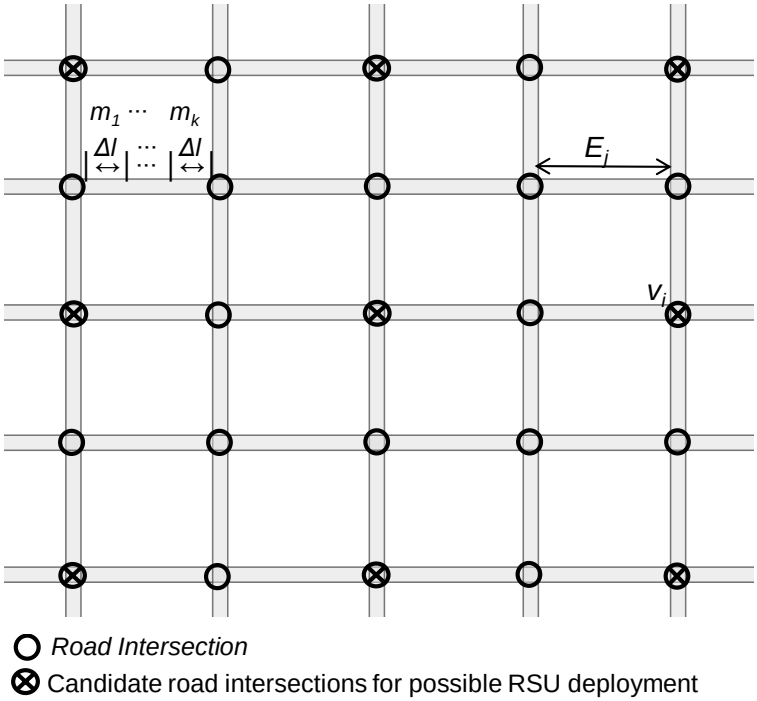
5.3.1.1 System Model

Consider the road network (shown in Figure 5.6(b)) as a graph with each intersection as a vertex and each road segment as an edge. $\boldsymbol{V}$ is set of all vertices, let $i \in \boldsymbol{V}$ *(or* $v_i \in \boldsymbol{V}$*).* $\boldsymbol{E}$ is set of all edges, let $j \in \boldsymbol{E}$. Each road segment is further divided into many smaller sub-segments (each of length Δl) $\boldsymbol{M}$ is the set of all such sub-segments in the complete road network, let $k \in \boldsymbol{M}$ (or $m_k \in \boldsymbol{M}$*).* For a sub-

segment $k \in \boldsymbol{M}$, let d_k be the vehicle density (vehicles/Km), f_k be the event/incident frequency (number of events happened in a given time – frequency of events) and s_k be the vehicle speed (Km/hr). The densities and speed on all sub-segments $k \in \boldsymbol{E}_j$ cannot always be the same because of different surface conditions (bumpy, slippery, etc), different gradient (steep climb, uphill, downhill, etc), different geometry (curving, straight, etc) and proximity to road signals or stop signs etc. Simplified event/incident distributions were considered to ease evaluation. The event/incident distribution functions for the road network of Figure 5.6(b) that will be evaluated in this paper are shown in Figure 5.7. Figure 5.7(a) shows a distribution where the likelihood of an event/incident changes from one road to another but is constant over one particular road. This can be the case when roads have different characteristics such as road widths, speed limits, vehicle densities, and neighborhoods. Figure 5.7(b) shows a distribution where the likelihood of an event/incident, in addition to changing from one road to another road, also changes over every road. This corresponds to the more realistic scenario where events/incidents are more likely to happen around intersections than in the road sub-segments that are far away from intersections.



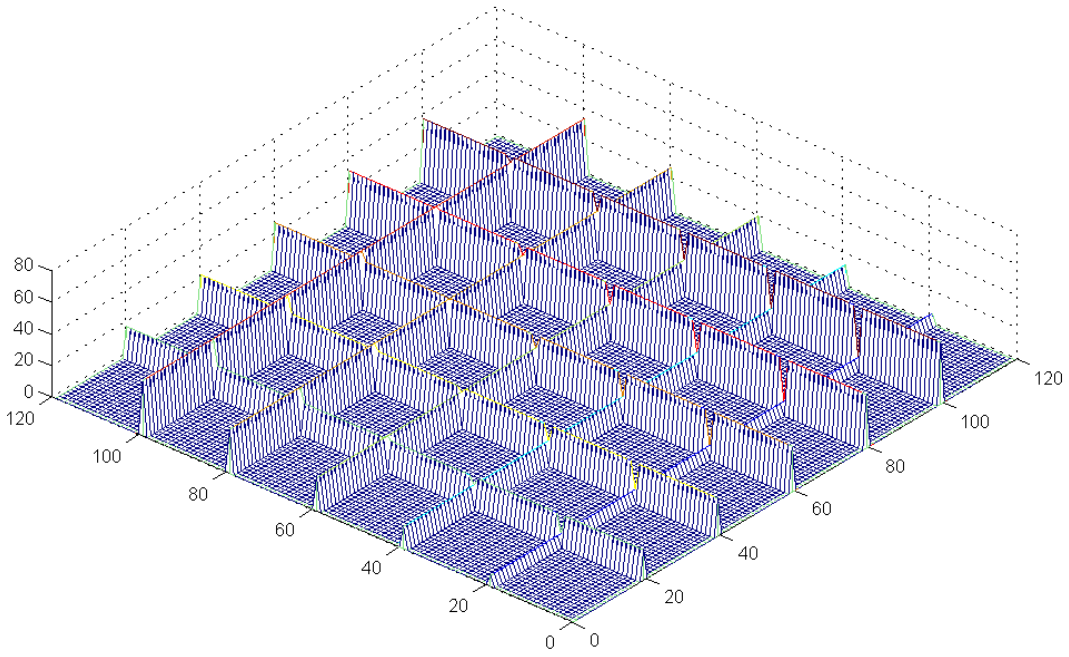
(a)



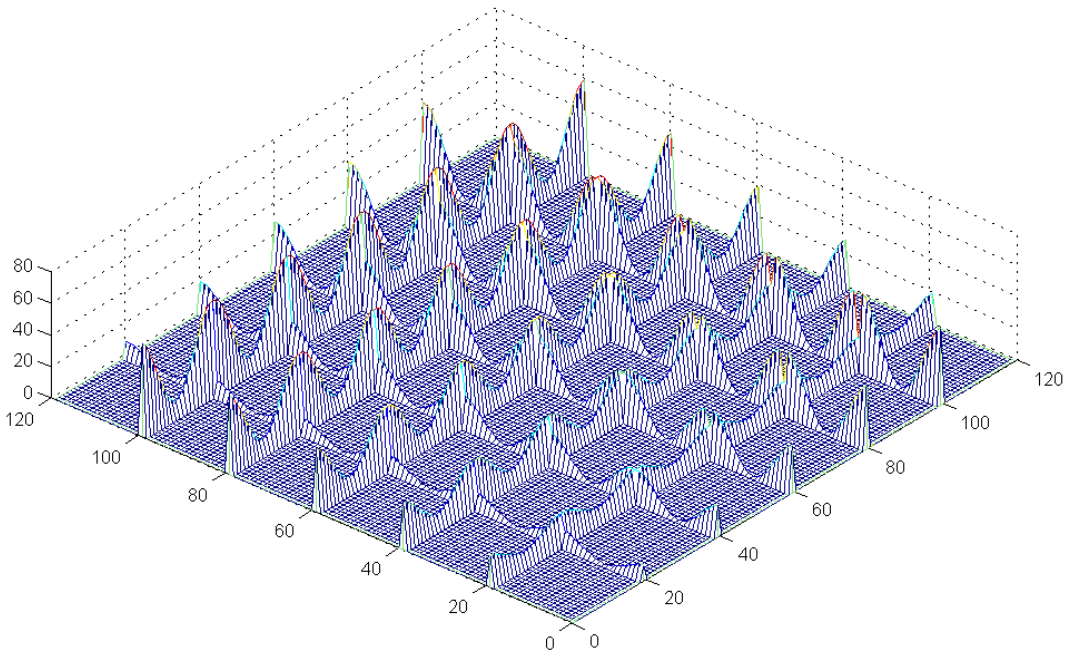
(b)

Figure 5.6: Urban environment. (a) Partial map of Miami, FL, USA. © OpenStreetMap contributors, CC-BY-SA (b) Grid-road network approximation of Figure 5.4(a).

If vehicles entering the region follow Poisson distribution then there will be $\lambda_k = s_k\ d_k$ vehicles entering the sub-segment $k \in \boldsymbol{M}$ *(or* $m_k \in \boldsymbol{M}$*)* per unit time. If y ($y \in \boldsymbol{M}$) is the location of an incident/event then a vehicle will reach the point of incident with an exponentially distributed time, with an average value of $1/\lambda_y$. If x ($x \in \boldsymbol{V}$) is the location of an RSU, then the reporting time, $t_{(x,y)}$ (time for a vehicle to report an incident happened at location y to an RSU at location x) will be the summation of the time for a vehicle to reach location y (t_y) and the time for the vehicle to reach x from y (t_{yx}), see Equation 5.4 and Figure 5.8.



(a)



(b)

Figure 5.7: Event/incident distribution functions: Relative frequency of events (*z axis*) at each segment (*x-y axes*). (a) Stair (b) Wave

$$t_{(x,y)} = t_y + t_{yx} = \frac{1}{(s_y d_y p_{yx})} + t_{yx} \tag{5.4}$$

where p_{yx} is the probability that a vehicle at $y \in \boldsymbol{M}$ will travel to $x \in \boldsymbol{V}$.

If there are more than one paths/routes from y to x, as will be the case in urban environment, then in Equation 5.4 t_{yx} should represent the average time taken by any vehicle at $y \in M$ to travel to $x \in V$ along all the possible routes. Its value is given by Equation 5.5.

$$t_{yx} = \sum_z t_{yxz} D_z \tag{5.5}$$

where z is the number of possible routes from $y \in \boldsymbol{M}$ to $x \in \boldsymbol{V}$, D_z is the fraction of vehicles travelling from $y \in \boldsymbol{M}$ to $x \in \boldsymbol{V}$ that use route z and t_{yxz} is the time for the vehicle to reach $x \in \boldsymbol{V}$ from $y \in \boldsymbol{M}$ using route z.

Let $\boldsymbol{N}$, $n \in \boldsymbol{N}$, be the set of sub-segments forming a route from $y \in \boldsymbol{M}$ to $x \in \boldsymbol{V}$. The average reporting for any event/incident along this route is given by Equation 5.6. If a route contains more than one sub-routes then we can use either the average travelling times (as given by Equation 5.5) or just the most direct/shortest route.

$$T_{(x,y)} = \frac{\sum_{n \in N} t_{(x|n)} f_n}{\sum_{n \in N} f_n} \tag{5.6}$$

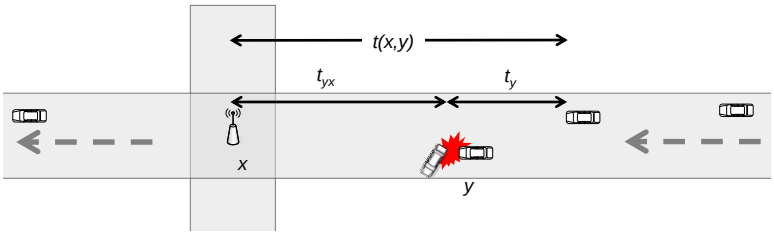


Figure 5.8: Reporting Time of an incident/event

5.3.1.2 Optimization Problem Modeling

Let C be the total number of sub-segments in a road network, i.e., $n(\boldsymbol{M})=C$ and R be the total number of intersections in a road network, i.e., $n(\boldsymbol{V})=R$ (we use notation $n(\boldsymbol{A})$ for number of elements in set $\boldsymbol{A}$). Each intersection is a candidate location for an RSU. If r is the desired number of RSUs, τ is the desired average reporting time and a is the desired fraction of coverage in the road

network ($m=aC$ is the number of covered sub-segments), then the two optimization problems can be stated as follows:

- **Minimize the average reporting time:** Minimize the average reporting time over each route (or an upper bound on the average reporting time over any route) such that at most r RSUs are placed among R candidate locations of set $\boldsymbol{V}$ and at least m out of C sub-segments of set $\boldsymbol{M}$ are covered by these RSUs.
- **Minimize the number of RSUs:** Minimize the number of RSUs placed among R candidate locations of set $\boldsymbol{V}$ such that at least m out of C sub-segments of set $\boldsymbol{M}$ are covered within τ average reporting time over each route (or an upper bound on the average reporting time over any route).

5.3.1.3 Problem Complexity

One possible optimization option is to exhaustively check all possible combinations to find an optimal solution. The number of possible combinations for optimization problem 1 and 2 are given by Equation 5.7 and Equation 5.8 respectively. The solutions that check all possible combinations to find an optimal solution increasingly become inefficient with the increase in size of area/region. For a 10Km x 10Km urban area with a grid-road topology, we may have a total road length of 100Km and 25 intersections. For a sub-segment size of 250m, we will have a total of 400 sub-segments. If we want to minimize the average reporting time for a total of r=5 RSUs and a=0.8 (80%) coverage, the number of possible combinations will be 1.15×10^{313}. And, if we want to minimize the number of RSUs (≤ 9) for some average reporting time and 80% coverage, the number of possible combinations will increase to 1.97×10^{397}.

$$\binom{R}{r}\binom{C}{m} r! \left\{ {m \atop r} \right\} \tag{5.7}$$

$$\sum_{r=1}^{R} \binom{R}{r}\binom{C}{m} r! \left\{ {m \atop r} \right\} \tag{5.8}$$

Where, $\binom{C}{m}$ is *combinations* that gives the number of subsets with size *m* when picking from a larger set of size *C,* and $\left\{ {m \atop r} \right\}$ is the *Stirling numbers of the second kind* that gives the number of ways to partition a set of *m* elements into *r* nonempty (and non-distinct) subsets [1]. It is given by Equation 5.9.

$$\left\{ {m \atop r} \right\} = \frac{1}{r!} \sum_{p=0}^{r} (-1)^p \binom{r}{p} (r-p)^m \tag{5.9}$$

5.3.2 Optimization Schemes

5.3.2.1 Binary Integer Programming (BIP) Optimization

The linear programming formalizations are not aware of the road topology so we need to relax the condition of the average reporting time that is defined over a single path/route to the average reporting time defined over entire region. The two performance metrics are not the same but the relaxation helps us to solve the optimization problem analytically using linear programming. It is important to note that averaging over entire region is more relaxed; it may include some routes whose average reporting time will be greater than the average reporting time over the entire region.

$\boldsymbol{M}$ is the set of all sub-segments in the road network, let ($k \in \boldsymbol{M}$); and $\boldsymbol{V}$ is the set of all intersections (candidate RSU locations), let ($i \in \boldsymbol{V}$). Let N_k be the number of incidents happening on any sub-segment $k \in \boldsymbol{M}$ and A_{ki} be the reporting time of an incident at sub-segment $k \in \boldsymbol{M}$ to an RSU $i \in \boldsymbol{V}$.

Let y_i and x_{ki} be two binary decision variable; such that, y_i equals to 1 if RSU $i \in V$ exists and 0 otherwise and x_{ki} equals to 1 if sub-segment $k \in M$ is covered by RSU $i \in V$ and 0 otherwise. The two optimization problem formulizations are as follows:

5.3.2.1.1 Minimize the Average Reporting Time (BIP-I)

The optimization goal is to minimize the average reporting time for a given number of RSUs and area coverage. As discussed earlier, we have relaxed the minimization of the average reporting time over each route constraint and replaced it with the average reporting time over the entire region. Specifically, here we minimize the total reporting time over the entire region. The binary integer programming formalization of this optimization problem is given in Figure 5.9.

Constraint (1) requires that the number of RSUs should be less than or equal to the desired value (r). Constraint (2) requires that each sub-segment is assigned to one or more than one RSUs, this ensures 100% coverage. Constraint (3) ensures that sub-segments are assigned to only those RSUs that are included in the solution. Constraints (2b and 6) replace constraint (2) if the required coverage is less than 100% coverage but equal to or greater than some coverage threshold (given by aC).

5.3.2.1.2 Minimize the Number of RSUs (BIP-II)

The optimization goal is to minimize the total number of RSUs for given average reporting time and area coverage. Average reporting time over each route has been relaxed to average reporting

time over entire region/area. The binary integer programming formalization of this problem is given in Figure 5.10.

Constraint (1) requires that the average reporting time over entire region is less than or equal to the average reporting time threshold (τ). Constraint (2) requires that each sub-segment is assigned to one or more than one RSUs, this ensures 100% coverage, i.e., all sub-segments are assigned to some RSU. Constraint (3) ensures that sub-segments are assigned to only those RSUs that are included in the solution. Constraints (2b and 6) replace constraint (2) if the required coverage is less than 100% coverage but equal to or greater than some coverage threshold (given by aC).

$$f(r) = min \quad \sum_{i \in V} \sum_{k \in M} N_k A_{ki} x_{ki}$$

$s.t.$

- $For \quad \alpha = 1 \qquad (100\%\ coverage)$

$$(1) \quad \sum_{i \in V} y_i \leq r$$

$$(2) \quad \sum_{i \in V} x_{ki} \geq 1 \qquad \forall\, k, k \in M$$

$$(3) \quad \sum_{k \in M} x_{ki} \leq C y_i \quad \forall\, i, i \in V$$

$$(4) \quad y_i \in \{0,1\} \qquad i \in V$$

$$(5) \quad x_{ki} \in \{0,1\} \qquad k \in M, i \in V$$

- $For\ \alpha < 1 \qquad (100\alpha\ percent\ coverage)$

$$(2b) \quad \sum_{i \in V} x_{ki} \leq 1 \qquad \forall\, k, k \in M$$

$$(6) \quad \sum_{i \in V} \sum_{k \in M} x_{ki} \geq \alpha C$$

Figure 5.9: Formulization BIP-I: Minimizing total reporting time *f(r)* for given *r* number of RSUs and *a* area coverage. For 100% coverage, the constraints are (1) to (5); for 100α percentage of coverage, the constraints are (1), (2b), (3), (4), (5) and (6).

$$f(\tau) = min \quad \sum_{i \in V} y_i$$

$s.t.$

- $For \quad \alpha = 1 \qquad (100\%\ coverage)$

$$(1) \quad \frac{\sum_{i \in V} \sum_{k \in M} N_k A_{ki} x_{ki}}{\sum_{i \in V} \sum_{k \in M} N_k x_{ki}} \leq \tau$$

$$(2) \quad \sum_{i \in V} x_{ki} \geq 1 \qquad \forall\, k, k \in M$$

$$(3) \quad \sum_{k \in M} x_{ki} \leq C y_i \quad \forall\, i, i \in V$$

$$(4) \quad y_i \in \{0, 1\} \qquad i \in V$$

$$(5) \quad x_{ki} \in \{0, 1\} \qquad k \in M, i \in V$$

- $For\ \alpha < 1 \qquad (100\alpha\ percent\ coverage)$

$$(2b) \quad \sum_{i \in V} x_{ki} \leq 1 \qquad \forall\, k, k \in M$$

$$(6) \quad \sum_{i \in V} \sum_{k \in M} x_{ki} \geq \alpha C$$

Figure 5.10: Formulization BIP-II: Minimizing total number of RSUs *f(τ)* for given *τ* reporting time (average reporting time over the entire region) and *a* area coverage. For 100% coverage, the constraints are (1) to (5); for 100α percentage of coverage, the constraints are (1), (2b), (3), (4), (5) and (6).

5.3.2.2 Balloon Expansion Heuristic (BEH) Optimization

In this optimization, each RSU and its coverage area is considered as a balloon that dynamically expands in a 2 dimensional space. A balloon's boundary represents the area covered by an RSU within a given average reporting time. The balloons are dynamically expanded as we gradually relax the average reporting time constraint till the desired percentage/fraction of area is covered by them.

The roads inside a balloon's boundary at any time include all the segments that can be covered by the RSU within some average reporting time via some route/path. The balloon expansion follows road network and the expansion is independent on each side, that is, if the RSU is located at an intersection then the balloon boundary on each of the four sides will expand independent of other three sides. The expansion depends on vehicle speed, vehicle density, event/incident distribution and probability of vehicles following a route. The segments, along a route, with high frequency of events/incidents will have more impact on computing the average reporting time than those with low frequency of events.

Figure 5.11 shows a road intersection where an RSU is located; A, B, C, and D is the balloon boundary for some average reporting time (τ). Initially, $|\boldsymbol{XA}|=|\boldsymbol{XB}|=|\boldsymbol{XC}|=|\boldsymbol{XD}|=0$, i.e., points {A, B, C, D} are located at X and $T_{(x,a)}=T_{(x,b)}=T_{(x,c)}=T_{(x,d)}=0$, where $T_{(x,y)}$ is the average reporting time along path $\boldsymbol{XY}$ including point Y. The balloon is then expanded independently on all four routes for some average reporting time (τ). The size of expansion on each route will vary depending on vehicle speed, vehicle density, incident/event distribution and probability of vehicles following a route. Figure 5.11 shows a balloon expansion where $|\boldsymbol{XA}|=|\boldsymbol{XD}|\leq|\boldsymbol{XC}|\leq|\boldsymbol{XB}|$.

Unrestricted expansions may form loops especially in urban environment. In order to avoid loops, we assume that the boundary expansion of an RSU is towards the direction away from the RSU; if the expansion encounters an intersection then it only continues in directions that are away from the RSU. Figure 5.12 gives the average reporting times for an RSU located at the center of an urban environment (Figure 5.6 (b)) for event/incident distributions given at Figure 5.7.

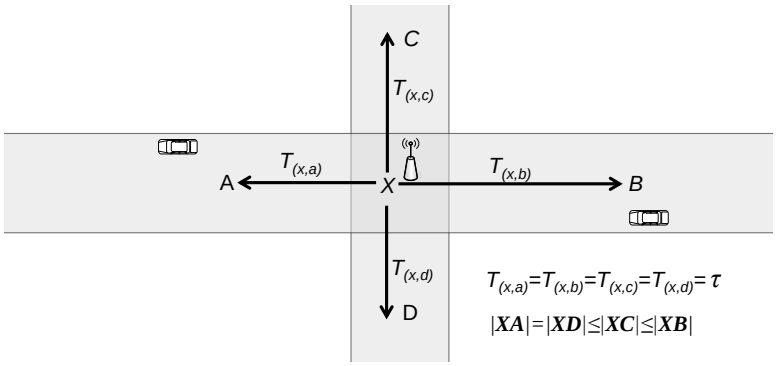
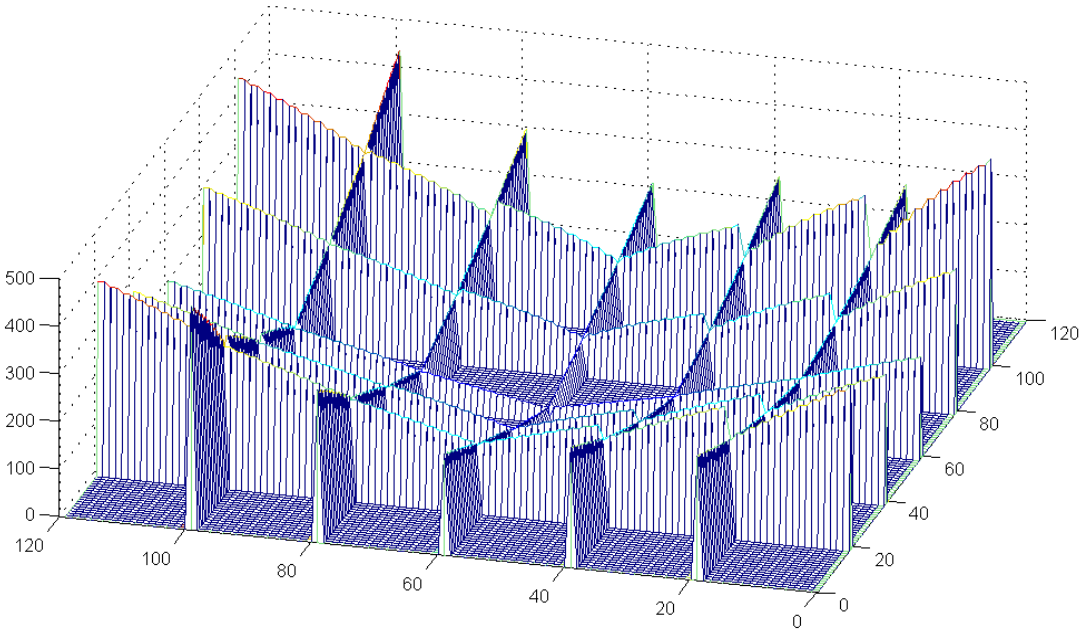
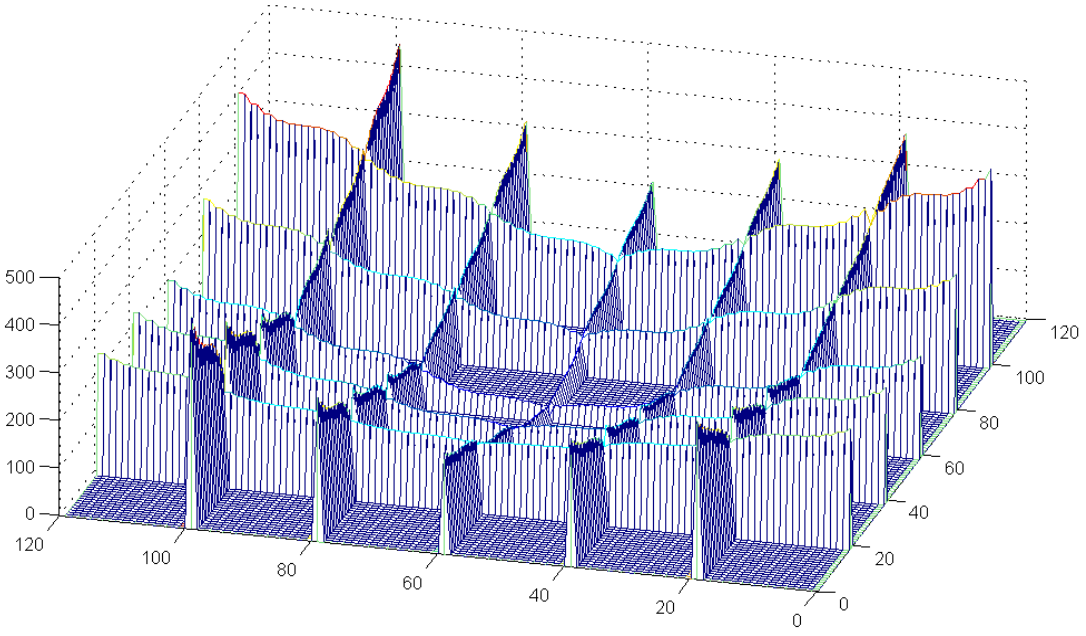


Figure 5.11: Balloon expansion: The expansion is independent along each direction and depends on vehicle speed, vehicle density, event/incident distribution and probability of vehicles following a route. $|XA|, |XD|, |XC|,$ and $|XB|$ gives the size of expansion towards A, B, C and D respectively for τ average reporting time over each route.

BEH optimization method, in general, starts with placing an RSU at each candidate location. The coverage of each RSU is then expanded on each side (along each route) for some value of average response time. The expansion continues till a sufficient number of sub-segments are covered by more than one neighboring RSUs (or the average reporting time threshold has reached). At this moment, the RSU with the least "impact factor" is removed (similar to the bursting of a balloon due to the too tight compression from neighboring balloons). The process repeats until the optimization objective is achieved. The *impact factor* of an RSU is the number of sub-segments that will not be covered if the RSU is removed; it is computed by subtracting the number of overlapped-sub-segments (sub-segments that are covered by this RSU and also by some other RSUs) from the number of sub-segments covered by this RSU.



(a)



(b)

Figure 5.12: Average reporting times, for different event/incident distributions, of urban environment given at Figure 5.7(b). (a) Stair (b) Wave

5.3.2.2.1 Minimize the Average Reporting Time (BEH-I)

The BEH algorithm for this optimization problem is given in Figure 5.13. The optimization objective is to minimize the average reporting time over each route (or the upper bound on average reporting time over any route) for given number of RSUs (r) and area coverage (αC). The method starts with placing an RSU at each candidate location (line 2), the average reporting time (of each route) is then iteratively incremented by a small value (line 5-8) until area coverage constraint is met (line 9-10). The impact factor of each RSU is calculated (line 11-13) and the one with the least impact factor (line 14-15) is removed provided the removal does not affect area coverage constraint (line 16-19); otherwise the average reporting time (of each route) will be further incremented (line 5-8). The process continues until the number of RSU constraint is met (line 4).

1. **begin**
2. $\boldsymbol{B} \leftarrow \boldsymbol{V}$
3. $\tau' \leftarrow 0$
4. **while** $n(\boldsymbol{B}) > r$ **do**
5. Increment τ' by a small value
6. **for** *each* $b \in \boldsymbol{B}$ **do**
7. Expand area coverage of b to $y\ s.t.\ T_{(b,y)} = \tau'$
8. **end**
9. Compute Q:
   combined area coverage by all RSUs in set $\boldsymbol{B}$
10. **if** $Q \geq \alpha C$ **then**
11. **for** *each* $b \in \boldsymbol{B}$ **do**
12. Compute Impact Factor $IF(b)$
13. **end**
14. Find $w \in \boldsymbol{B}\ \ s.t. min(IF) = IF(w)$
15. $\boldsymbol{B}' \leftarrow \{w\}$
16. Compute Q:
    combined area coverage by all RSUs in set $\{\boldsymbol{B} - \boldsymbol{B}'\}$
17. **if** $Q \geq \alpha C$ **then**
18. $\boldsymbol{B} \leftarrow \boldsymbol{B} - \boldsymbol{B}'$
19. **end**
20. **end**
21. **end**
22. **Return** ($\boldsymbol{B}$)
23. **end**

Figure 5.13: Algorithm BEH-I: Minimizing average reporting time over each route (i.e., the upper bound on average reporting time over any route) for given number of RSUs and area coverage. After the algorithm finishes, τ' gives the upper bound on the average reporting time over any route.

5.3.2.2.2 Minimize the number of RSUs (BEH-II)

BEH algorithm for this optimization problem is given in Figure 5.14. The optimization objective is to minimize the number of RSUs for a given average reporting time over each route (or an upper bound on the average reporting time over any route) (τ) and area coverage (αC). The method starts with placing an RSU at each candidate location (line 2). The coverage balloons of all the RSUs are then expanded till the average response time is equal to the desired value (line 4-6). The combined coverage is checked against desired coverage (line 7-9). The impact factor of each RSU is calculated (line 13-15) and the one with the least impact factor (line 16-17) is removed provided that the removal does not affect area coverage constraint (line 18-19, 11-12). Impact factors are then recalculated and a fresh sorted list is generated. The process stops when removing an RSU will dissatisfy the area coverage constraint (line 11).

1. **begin**
2. $\boldsymbol{B} \leftarrow \boldsymbol{V}$
3. $\boldsymbol{B'} \leftarrow \emptyset$
4. **for** *each* $b \in \boldsymbol{B}$ **do**
5. Expand area coverage of b to y $s.t.$ $T_{(b,y)} = \tau$
6. **end**
7. Compute Q:
   combined area coverage by all RSUs in set $\boldsymbol{B}$
8. **if** $Q < \alpha C$ **then**
9. **Return** ("Area coverage constraint cannot be met")
10. **else**
11. **while** $Q \geq \alpha C$ **do**
12. $\mathbf{B} \leftarrow \mathbf{B} - \mathbf{B'}$
13. **for** *each* $b \in \boldsymbol{B}$ **do**
14. Compute Impact Factor $IF(b)$
15. **end**
16. Find $w \in \boldsymbol{B}$ $s.t.$ $min(IF) = IF(w)$
17. $\boldsymbol{B}' \leftarrow \{w\}$
18. Compute Q:
    combined area coverage by all RSUs in *set* $\{\boldsymbol{B} - \boldsymbol{B}'\}$
19. **end**
20. **end**
21. **Return** ($\boldsymbol{B}$)
22. **end**

Figure 5.14: Algorithm BEH-II: Minimizing the total number of RSUs for given average reporting time over each route (i.e., the upper bound on average reporting time over any route) and area coverage

5.3.3 Simulation Results and Discussion

5.3.3.1 Simulation Setup

The simulation is based on an urban region with five vertical and five horizontal roads, as shown in Figure 5.6(b). The region is 3 Km x 3 Km, with a total road length of 30 Km. The sub-segment size is 250m, resulting in a total of 120 sub-segments. There are a total of 25 intersections; in order to reduce problem complexity for BIP methods (explained earlier), only 9 out of the 25 intersections are considered as candidate locations for RSUs (refer Figure 5.6(b)). Two different incident/event distributions are defined, as shown in Figure 5.7. Figure 5.7(a) shows a distribution where the likelihood of an event/incident changes from one road to another but is constant over one particular road. This can be the case when roads have different characteristics such as road widths, speed limits, vehicle densities, and neighborhoods. Figure 5.7(b) shows a distribution where the likelihood of an event/incident, in addition to changing from one road to another road, also changes over every road. This corresponds to the more realistic scenario where events/incidents are more likely to happen around intersections than in the road sub-segments that are far away from intersections. Vehicles entering the region follow Poisson distribution, with $\lambda=SD$ vehicles entering any sub-segment per unit time. A constant vehicle density of D = 4vehicle/Km and a constant speed of S = 50Km/hr is assumed for this simulation. The probability that a vehicle at a point of event/incident will travel to a particular RSU is considered to be inversely proportional to the number of intersections (or routes) between the vehicle and the RSU. The most direct and shortest path is used to calculate the reporting time of an event/incident to a particular RSU; only the vehicles following that route are considered in computing the average reporting time and the contribution by the rest of vehicles for reporting the event/incident is ignored (which, if considered,

may improve the event/incident reporting likelihood). It is important to note that in real scenarios/applications vehicle traces can be used to generate all these statistics; the statistics based on vehicle traces are generally reliable as daily traffic patterns are often repeated.

In order to study how well our proposed optimization methods could achieve, enumeration method was used to exhaustively search for the true optimal solution. As discussed earlier, enumeration method increasingly becomes inefficient with the increase in size of area/region. In order to reduce the number of combinations to be checked to find an optimal solution, so that we can actually finish the enumerating operation using personal computers, the average reporting time over each route constraint was relaxed and was replaced with the average reporting time over the entire region. With this relaxation, we can simply consider a segment to be covered by one RSU (out of all the currently considered RSUs) that has minimum reporting time from that segment instead of considering all of RSUs.

The specifications of system used for simulation are: *Processor* - Intel® Core™2 Quad CPU Q6700 @ 2.66 GHz, *RAM* - 4 GB, *Hard Disk* - 232 GB (200 GB free), and *Operating System* - Windows 7 Enterprise 64-bit.

5.3.3.2 Results

5.3.3.2.1 Enumeration/Exhaustive Search

The minimum average reporting time (over the entire region) for different number of RSUs, using enumeration/exhaustive search for different event/incident distributions of urban environment given at Figure 5.6(b), are shown in Figure 5.15. The result covers all possible number of RSUs, refer Equation 5.8, therefore it can be used to find the minimum number of RSUs required for a given average reporting time (over the entire region). For example, as shown in Figure 5.15, a minimum of 5 RSUs will be required for an average reporting time of 150 sec.

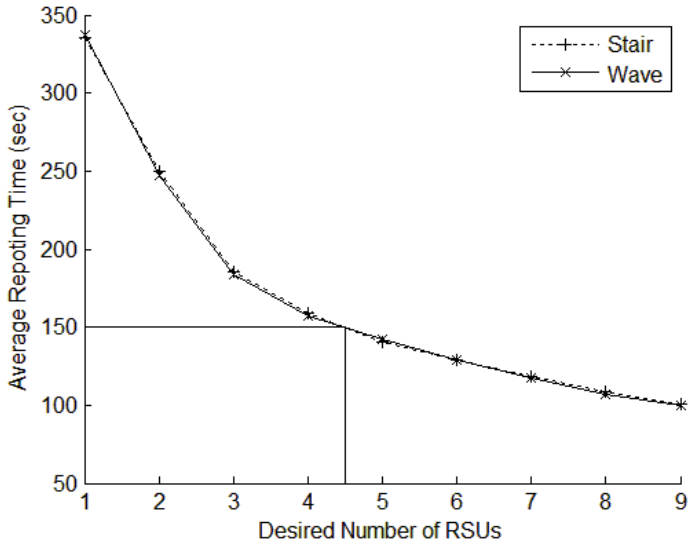
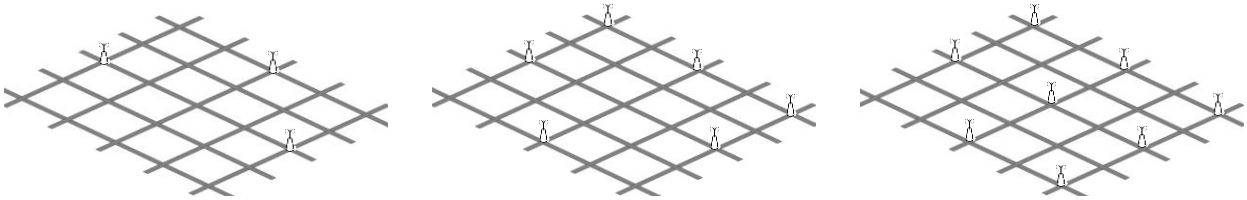


Figure 5.15: Minimum average reporting time (over the entire region) for different number of RSUs using enumeration/ exhaustive search for different event/incident distributions of urban environment given at Figure 5.6(b).

5.3.3.2.2 Binary Integer Programming (BIP) Optimization

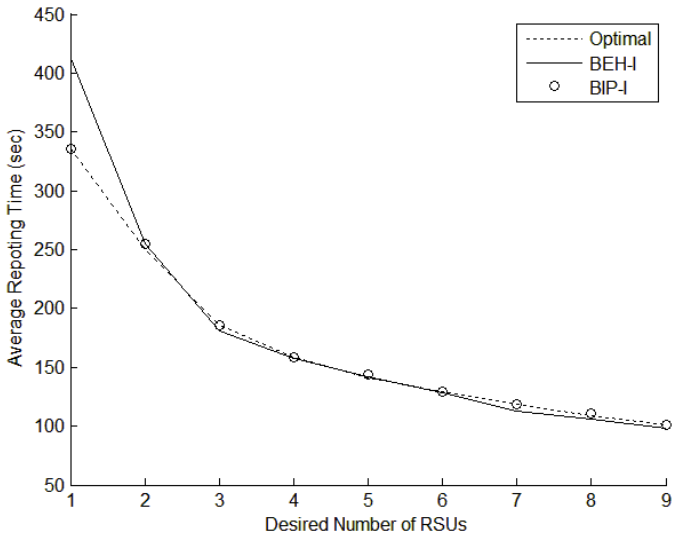
The optimal placements of RSUs for minimizing the total reporting time for different numbers of RSUs (BIP-I) are shown in Figure 5.16.



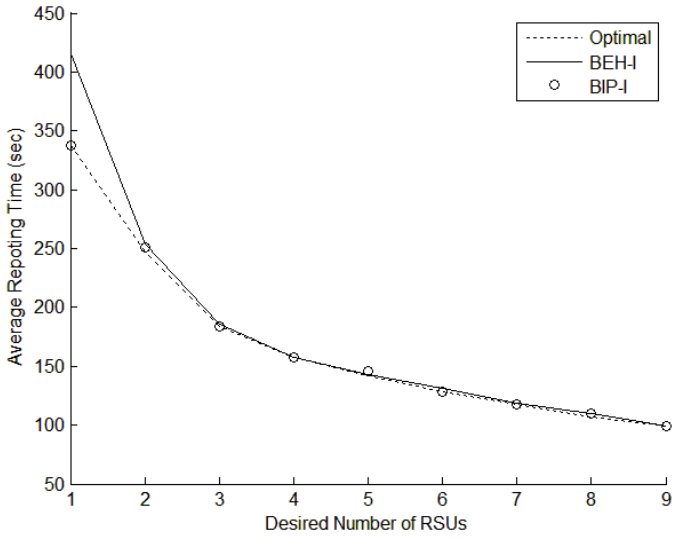
(a) (b) (c)

Figure 5.16: Optimal RSU placements using BIP-I (minimizing total reporting time for given number of RSUs and area coverage): (a) Number of RSUs = 3 (b) Number of RSUs = 6 (c) Number of RSUs = 8

The minimum average reporting time (over the entire region) for different number of RSUs, using *BIP-I* for different event/incident distributions of urban environment given at Figure 5.6, is shown in Figure 5.17. The minimum average reporting time over the entire region of *BIP-I* is the same as that of *enumeration/exhaustive search* solution.



(a)



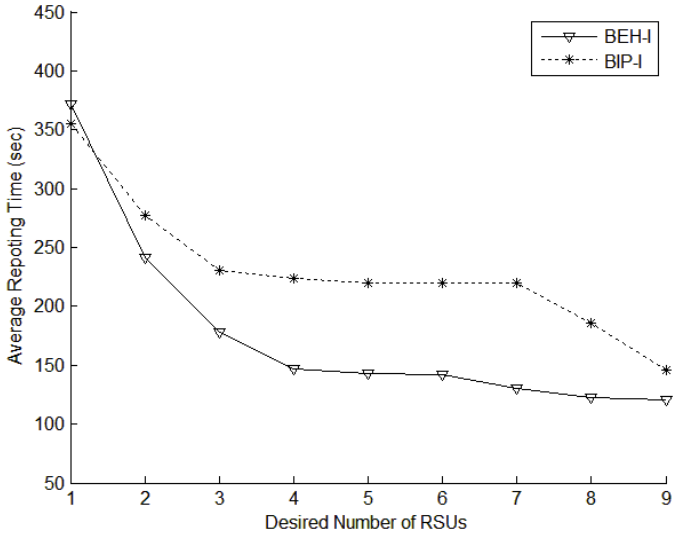
(b)

Figure 5.17: Minimum average reporting time (over the entire region) for different number of RSUs and different event/incident distributions of urban environment given at Fig. 1(b). (a) Stair (b) Wave

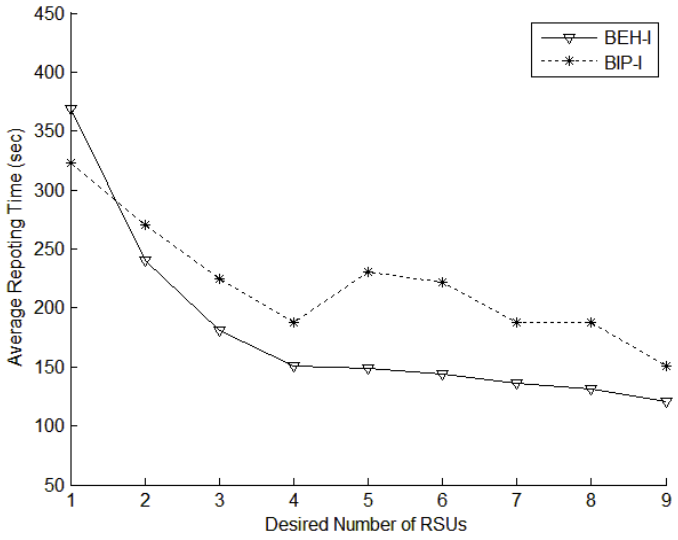
The minimum average reporting time over each route (or an upper bound on the average reporting time over any route) for different number of RSUs, corresponding to optimal solutions of *BIP-I* for

different event/incident distributions of urban environment given at Figure 5.6, are shown in Figure 5.18. The minimum average reporting time over each path of *BIP-I* is higher than that of *BEH-I*. The execution times for *BIP-I* is given in Figure 5.21.

Optimal placement of RSUs for minimizing the total number of RSUs (*BIP-II*) did not converge to a feasible solution within reasonable time (≤ 4 days).



(a)



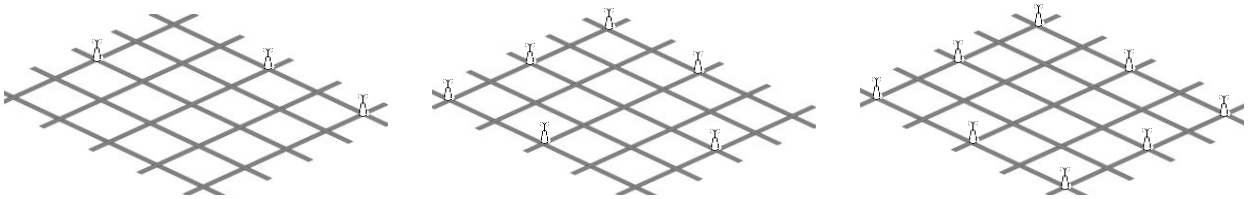
(b)

Figure 5.18: Minimum average reporting time over each route (or an upper bound on the average reporting time over any route) for different event/incident distributions of urban environment given at Fig. 1(b). (a) Stair (b) Wave

5.3.3.2.3 Balloon Expansion Heuristic (BEH) Optimization

The optimal placements of RSUs for minimizing the average response time over each route (or an upper bound on the average reporting time over any route) for different numbers of RSUs (BEH-I) are shown in Figure 5.19.

The minimum average reporting time over each route (or an upper bound on the average reporting time over any route) for different number of RSUs, using *BEH-I* for different event/incident distributions of urban environment given at Figure 5.6, are shown in Figure 5.18. The minimum average reporting time over each path achieved by *BEH-I* is better than that of *BIP-I*.

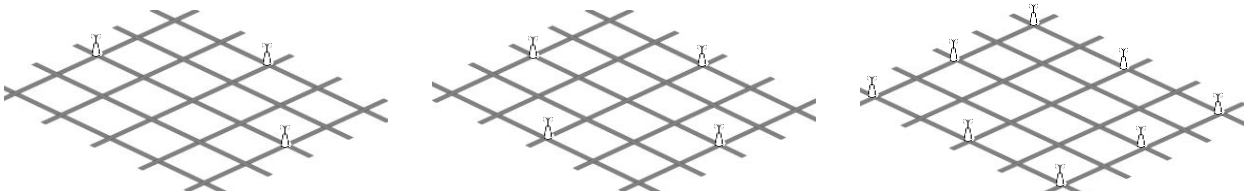


(a) (b) (c)

Figure 5.19: Optimal RSU placements using BEH-I (minimizing average reporting time over each route, or an upper bound on the average reporting time over any route, for given number of RSUs and area coverage): (a) Number of RSUs = 3 (b) Number of RSUs = 6 (c) Number of RSUs = 8

The minimum average reporting time (over the entire region) for different number of RSUs, corresponding to optimal solutions of *BEH-I* for different event/incident distributions of urban environment given at Figure 5.6, are shown in Figure 5.17. The minimum average reporting time over the entire region achieved by *BEH-I* closely follows that of enumeration/exhaustive search.

The optimal placement of RSUs for minimizing number of RSUs for different reporting times *(BEH-II)* are shown in Figure 5.20. The execution times for both the *BEH* algorithms are given in Figure 5.21 and 5.22.



(a) (b) (c)

Figure 5.20: Optimal RSU placements using BEH-II (minimizing total number of RSUs for given average reporting time over each route and area coverage) : (a) Average Reporting time ≤ 180 secs (b) Average Reporting time ≤ 150 secs (c) Average Reporting time ≤ 130 secs

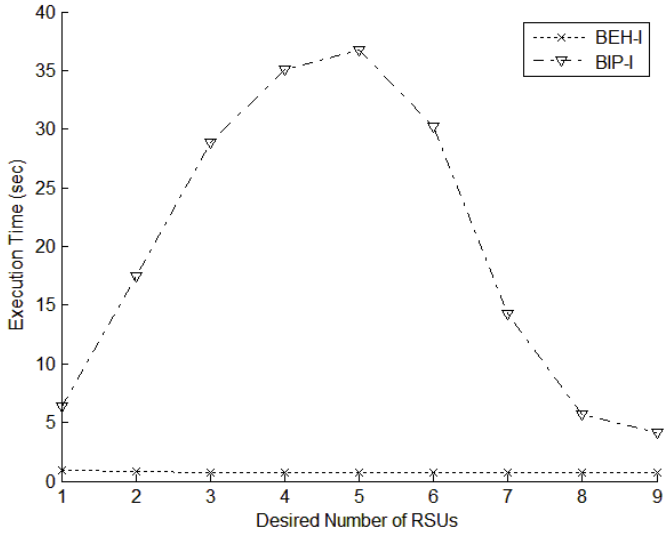


Figure 5.21: Execution times for BEH-I and BIP-1 for different number of desired RSUs

5.3.3.3 Discussion

BIP-I successfully produced optimal solutions. The minimum average reporting time over the entire region is the same as that of enumeration/exhaustive search (Figure 5.17). However, the minimum average reporting time over each path is higher than that of BEH-I (Figure 5.18).

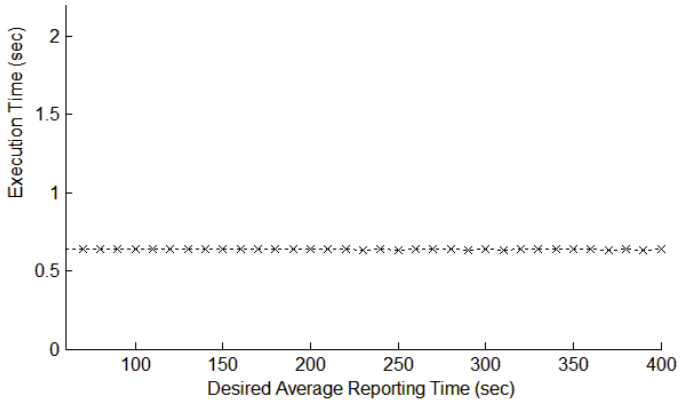


Figure 5.22: Execution times for BEH-II for different average reporting times

BIP-II did not produce feasible solution within reasonable time. BIP optimizations use branch and bound algorithm to solve the problems; the branch and bound algorithm uses binary search tree whose size grows tremendously with the size increase of a problem. The failure of *BIP-II* to converge within reasonable time may be due to reason that, in worst cases, branch and bound algorithm searches all possible combinations to find the best solution [2] and we have already shown in section II that the number of possible combinations for this problem is very large ($>10^{122}$).

One possible solution is to iteratively use *BIP-I* to solve *BIP-II*. The BIP-II optimization problem can be restated as: Find the smallest number of RSUs satisfying the average reporting time constraint (τ). A simple way is to repeatedly use *BIP-I,* with decreasing numbers of RSUs, to find the smallest number of RSUs that gives a minimum reporting time satisfying the timing constraint. The number of *BIP-I* computations can be reduced by employing various search methods such as binary search. The binary-search based algorithm is listed in the Figure 5.23.

1. **begin**
2. $Low \leftarrow 1$
3. $High \leftarrow R$
4. $Final\_assignment \leftarrow \emptyset$
5. **while** $Low \leq High$ **do**
6. $r' \leftarrow \left\lfloor \frac{Low+High}{2} \right\rfloor$
7. $Current\_assignment \leftarrow f(r')$
8. Compute average reporting time τ' for $Current\_assignment$
9. **if** $\tau' \leq \tau$ **then**
10. $High \leftarrow r' - 1$
11. $Final\_assignment \leftarrow Current\_assignment$
12. **else**
13. $Low \leftarrow r' + 1$
14. **end**
15. **end**
16. **Return** $(Final\_assignment)$
17. **end**

Figure 5.23: Algorithm BIP-II.I (Using Binary search and BIP-I): Minimizing total number of RSUs for a given average reporting time constraint and area coverage

BEH algorithms incorporate the knowledge of road topology to find the optimal solution. Both the *BEH algorithms* successfully produced optimal solutions. The execution times for both the *BEH* methods are much less than that of *BIP-I* method (refer Figure 5.21 and 5.22).

In addition, the minimum average reporting time over each path achieved by BEH is better than that of BIP-I (Figure 5.18). The minimum average reporting time over the entire region achieved by BEH closely follows that of enumeration/exhaustive search (Figure 5.17).

BEH-I removes RSUs as it increments the average reporting time. The selection of an RSU to be removed, at any stage, may vary with the size of average reporting time increment. A smaller increment size gives higher resolution on minimum average reporting time, whereas a larger increment size may result in more global optimization. *BEH-I* removes RSUs after incrementing the average reporting time to the desired threshold. It then greedily removes RSUs with the least impact factors; this RSU removal can also be modeled as a knapsack problem [14], in future work, for more fine grained solutions.

The RSU removal sequence, in both the *BEH algorithms*, may later be utilized for incrementally deploying RSUs. *BEH algorithms* provide the optimal RSU placement for a given (or minimum) number of RSUs, later if we want to add more RSUs to the region, then they can be deployed at locations that were removed last. *BEH-I* can also be used to find marginal improvement in average reporting time for each added RSUs.

BEH algorithms use average reporting time over a path as a timing constraint. Average reporting time over a path is more useful metric then average reporting time over entire region; it guarantees that on average an event/incident will be reported within the timing constraint whereas the average reporting time over entire region cannot guarantee this.

5.4 Conclusion

We have presented an optimization schemes for two different environments: along highways and in urban areas. For highways, the optimization aims to minimize the average reporting time of an

event/incident by a vehicle to a nearby RSU. Our optimization scheme is based on balloon expansion analogy, where the expansion in each direction is related to vehicle speed, vehicle density and likelihood of incidents/events. We have shown that our balloon optimization scheme performs equally well as the exhaustive optimization scheme. For urban areas, we have presented two optimization methods: Binary Integer Programming (BIP) method and Balloon Expansion Heuristic (BEH) method. Both optimization methods were used to solve two optimization problems: minimizing the average reporting time and minimizing the number of required RSUs. We have shown that the novel BEH method is more versatile and can be used to solve both the optimization problems without any further relaxations.

5.5 References

[1] M. Abramowitz, and I.A. Stegun, (Eds.). "Stirling Numbers of the Second Kind." §24.1.4 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 824-825, 1972.

[2] J. Clausen, "Branch and Bound Algorithms—Principles and Examples," Dept. Comput. Sci., Univ. Copenhagen, [Online]. Available: http://www.diku.dk/OLD/undervisning/ 2003e/datV-optimer/ JensClausenNoter.pdf.

[3] J. Lee and C. Kim, "A roadside unit placement scheme for Vehicular Telematics networks", in AST'10, LNCS vol. 6059, pp. 196-202, 2010.

[4] P. Li, X. Huang, Y. Fang and P. Lin, " Optimal placement of gateways in Vehicular Networks", in IEEE Transactions on Vehiclular Technology 2007, Vol. 56/ 6, pt 1, pp. 3421-3430, 2007.

[5] W. Zhao, Y. Chen, M. Ammar, M. Corner, B. Levine and E. Zegura, "Capacity enhancement using Throwboxes in DTNs". In IEEE MASS'06, 2006.

[6] C. Lochert, B. Scheuermann, C. Wewetzer, A. Luebke and M. Mauve, "Data aggregation and roadside unit placement for a VANET traffic information system. In ACM VANET'08, pp. 58—65, Sep. 2008.

[7] Y. Sun, X. Lin, R. Lu, X. Shen, J. Su, "Roadside Units Deployment for Efficient Short-time Certificate Updating in VANETs", In IEEE ICC'10, 2010.

[8] M. Fiore, J. Barcelo-Ordinas,"Cooperative download in urban vehicular networks", in IEEE MASS '09, 2009.

[9] O. Trullols, M. Fiore, C. Casetti, C.F. Chiasserini, J.M. Barcelo Ordinas, "Planning roadside infrastructure for information dissemination in intelligent transportation systems", in Computer Communications, Vol. 33/ 4, pp 432-442, March 2010.

[10] F. Malandrino, C. Casetti, C. Chiasserini, M. Fiore , "Content downloading in vehicular networks: What really matters," in INFOCOM'11, April 2011.

[11] Z. Zheng, Z. Lu, P. Sinha, S. Kumar, "Maximizing the Contact Opportunity for Vehicular Internet Access," in INFOCOM'10, March 2010.

[12] M. L. Brandeau and S. S. Chiu, "An overview of representative problems in location research", Management Science, 35(6):645–674, 1989.

[13] R. Francis and J. White, "Facility Layout and Location: An Analytical Approach", Englewood Cliffs, NJ: Prentice-Hall, 1974.

[14] Martello, Silvano; Toth, Paolo, "Knapsack Problems: Algorithms and Computer Interpretations", Wiley-Interscience, 1990.

CHAPTER 6 CONCLUSION

Need of ubiquitous connectivity has brought connectivity to vehicles, resulting in formation of VANET. A lot of work has been done in this area and various applications, protocols and standards have been developed. Most of the applications and services assume a mature VANET with pervasive roadside infrastructure and large number of smart vehicles. But, during the initial deployment stages of VANET there will be very scarce roadside infrastructure and very limited number of smart vehicles on the road. Therefore, in order to bootstrap VANET and make VANET an attractive investment for both the service providers and consumers there is a need to develop various solutions that do not depend on pervasive roadside infrastructure or a large number of vehicles. We have proposed various solutions that are viable even with very scarce roadside infrastructure and very limited number of smart vehicles. These solutions are economical, scalable, and deployable and will help in stimulating VANET activity and its adoption.

The contributions of this dissertation are summarized below.

- Design of a VANET architecture that does not need expensive roadside infrastructure or large number of smart vehicles, can provide services with limited number of heterogeneous roadside units and smart vehicles with varying capabilities, is scalable and deployable.
- Provision of backend connectivity to Internet to smart vehicles without requiring pervasive roadside infrastructure or large number of smart vehicles, especially in rural areas and along

highways. Satellite down-only link is incorporated to provide connectivity during long intervals between RSUs.

- Design of security architecture that does not depend on pervasive roadside infrastructure or a fully connected V2V network. The architecture is economical, scalable and deployable. Two architectures have been proposed: a service oriented and a general purpose. Both architectures fulfill the security requirements.
- Optimal placement of limited number of RSUs within a given area to provide best possible service to smart vehicles. The optimal placement solution covers both environments: the urban areas and the highways. A novel heuristic optimization solution has been proposed that performs near optimal results without needing extensive resources as that required by exhaustive search optimization solutions.