Towards Efficient Certificate Revocation Status Validation in Vehicular Ad Hoc Networks with Data Mining

By

Qingwei Zhang

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements

For the M.A.Sc. degree in Electrical and Computer Engineering

School of Electrical Engineering and Computer Science

Faculty of Engineering

University of Ottawa

Abstract

Vehicular Ad hoc Networks (VANETs) are emerging as a promising approach to improving traffic safety and providing a wide range of wireless applications for drivers and passengers. To perform reliable and trusted vehicular communications, one prerequisite is to ensure a peer vehicle's credibility by means of digital certificates validation from messages that are sent out by other vehicles. However, in vehicular communication systems, certificates validation is more time consuming than in traditional networks, due to the fact that each vehicle receives a large number of messages in a short period of time. Another issue that needs to be addressed is the unsuccessful delivery of information between vehicles and other entities on the road as a result of their high mobility rate. For these reasons, we need new solutions to accelerate the process of certificates validation. In this thesis, we propose a certificate revocation status validation scheme using the concept of clustering; based on data mining practices, which can meet the aforementioned requirements. We employ the technique of k-means clustering to boost the efficiency of certificates validation, thereby enhancing the security of a vehicular ad hoc network. Additionally, a comprehensive analysis of the security of the proposed scheme is presented. The analytical results demonstrate that this scheme can effectively improve the validation of certificates and thus secure the vehicular communication in vehicular networks.

List of Publications

The following publications, which include both accepted and published research articles, are relevant to the topic of this thesis and have been co-authored by Qingwei Zhang:

Conference Publication

Qingwei Zhang, Mohammed Almulla, Yonglin Ren, and Azzedine Boukerche, "An Efficient Certificate Revocation Validation Scheme with k-Means Clustering for Vehicular Ad hoc Networks," *2012 IEEE Symposium on Computers and Communication (ISCC)*, pp. 0862-0867, July 1 - 4, 2012.

Journal Publication

Mohammed Almulla, **Qingwei Zhang**, Yonglin Ren, and Azzedine Boukerche, "An Efficient k-Means Authentication Scheme for certificate revocation validation in Vehicular Ad hoc Networks," accepted in *Wireless Communications and Mobile Computing (WCMC),* John Wiley & Sons, Ltd., 2012.

Acknowledgements

It is a pleasure to thank all those who made this thesis possible.

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Azzedine Boukerche, and to Dr. Mohammed A. Almulla for their continuous support throughout my Master's study and research, and for their inspiration, motivation, enthusiasm, and immense knowledge.

The numerous discussions we had about the ideas and protocols proposed in this thesis, and their guidance helped me to go through much of the research and the writing of technical papers and this thesis. I could not have imagined having a better supervisor and mentor for my Master's study.

I also wish to express my warm and sincere thanks to a dear friend and PARADISE Student Yonglin Ren. His friendship, patience, encouragement and his help have provided me with a good basis to carry out this thesis work.

I am grateful to my many other fellow labmates in the DIVA Group at University of Ottawa for all the fun we have had in the last two years. This includes Renfei Wang, for the sleepless nights we were working together before deadlines; Amar Farouk Merah, for the stimulating discussions, and lot of amazing ideas; Maram Bani Younes, for her guidance in using Network Simulatior-2. I wish also to thank in addition both Robson De Grande and Cristiano Gato De Rezende for making my stay at PARADISE Research Laboratory fun, they certainly deserve special mention.

Last but not least, I would like to thank my family, in particular my parents Mr. Baixin Zhang and Mrs. Huiqiong Zeng, for giving birth to me in the first place and supporting me spiritually throughout my life.

Dedication

This thesis is dedicated to my FATHER, who enlighten me to the fact that no victory comes without a price and there are no accidents.

It is also dedicated to my MOTHER, who taught me that even the largest task can be accomplished if it is done one step at a time.

Finally, I lovingly dedicate this thesis to two families: the Tabicas and the Dorions. Francisco, Sylvie and their cute daughter, Amelia, all of you support me each step of my way. Andre and Suzanne, you offered me unconditional love and had my health in your heart throughout the life of my stay in Canada.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

There has been growing interest in developing secure and convenient driving conditions in the past decades. Vehicular Ad hoc Networks (VANETs) play an important role in wireless communication amongs vehicles, which emphasizes drivers' safety and the use of assistance applications[2][3].

In VANETs, vehicular communication allows vehicles and infrastruactures to communicate with each other over single or multiple hops. VANETs consist of several main components, the function of which are discussed below:

- *Vehicles (*also referred to as *nodes)*: in VANETs, vehicles act as either terminals or routers, sending and forwarding messages to the destination.
- *Certificate Authorities (CAs)*: the Certificate Authorities' main goal is to help in securing access of trusted information. This is an indispensable part in VANETs' structure. Serving as a completely trusted party, it may be a road administrate bureau of government that is responsible for maintaining a vehicle's information or it can be a company running a range of security services. Furthur details will be provided regarding CAs in subsequent chapters.
- *On Board Units (OBUs)*: an On Board Unit will be equipped in each vehicle: This is a physical electrical device that is able to transmit messages regarding

the direction, position, speed, etc. OBUs are also able to verify whether or not incoming messages are received from valid nodes.

- *Road Side Units (RSUs)*: Similar to OBUs, RSUs are fixed electrical hardware devices that are responsible for collecting personal information of passing vehicles and providing driver-assistance applications with necessary support, such as the navigation or warning message dissemination. Moreover, the network access devices in RSUs are in charge of keeping vehicles online. They are commonly deployed in intersection zones and function as a gateway that enables vehicles to keep connection with the Internet.

As a self-organized system, a VANET has its intrinsic and unique characteristics, which not only bring drivers and road administrators a wide set of on-road assistance applications, but also can potentially cause severe consequences and be subject to attacks. Therefore, the security of vehicular networks is critical and indispensable. It is also inevitable that latency requirements exist for applications and services in vehicular communication since different data transmitted over VANETs should be classified into different critical levels, namely safety or Non-safety, for different consideration. Further, based on the time-sensitivity of *Quality of Service (QoS)* applications, they can be categorized as Life-critical (such as Emergency Braking Warning), Real-time (for example, video and audio, etc.) and Non-real-time (for instance, surfing internet, checking email etc.). Life-critical messages must be sent with a threshold that indicates the restriction of latency, otherwise they might arrive at the destination late and would thus be considered useless. In contrast to Life-critical messages, the other two classifications are more tolerant to latency. Table 1.1 shows some examples of the safety and non-safety applications with different delay bounds and broadcasting interval requirements, envisioned for the The *Dedicated Short-Range Communication (DSRC)* spectrum [4].

In order to establish a reliable vehicular communication environment, the guarantee of a node's credibility is required. Usually, authentication and digital certificates act as

Table 1.1: An example of DSRC applications and requirements.

Application	Allowable Latency (ms)	Network Traffic	Priority
Intersection collision Warning	~100	Event	Safety of Life
Emergency Braking Warning	~100	Event	Safety of Life
Cooperative Collision Warning	~100	Periodic	Safety of Life
Work Zone Warning	~1000	Periodic	Safety
Transit Vehicle Signal Priority	~1000	Event	Safety
Toll collection	~50	Event	Non-Safety
Service Announcements	~500	Periodic	Non-Safety

the major tools used to validate the identification of each communicating entity. One entity's certificate can be validated by checking its digital certificates and an efficient authentication methodology must ensure the trustworthiness of a message with promptness by verifying its sender's certificate revocation status.

In fact, the promptness of validation is much more important for VANETs when compared to conventional networks because it is not unusual that a single vehicle will receives a large number of messages in a short amount of time, for example many vehicles stuck in a traffic congestion. Moreover, keeping connections live between different entities can be extremely challenging to achieve due to the high velocity of moving vehicles and the increased distance between these vehicles as they move in different directions. Thus, it is necessary to find an efficient scheme to expedite the certificates validation process.

In this thesis, we propose a novel certificates validation scheme that adopts the concept of clustering based on the data mining model.

1.1 Motivation

Since providing vehicles an easy and convenient method to share the latest traffic information and numerous types of data is increasingly important, VANETs have become a meaningful application of *Mobile Ad hoc Networks (MANET)* that have demonstrated success achievement and excellence. According to the general specifications of vehicular communications, each vehicle on the road should share messages that contain its current position, speed and other communication information to other vehicles on a periodic basis. Therefore, it is sometimes possible for each vehicle to receive a large number of messages that are shared by other peers within a certain period.

Upon receiving messages from neighboring vehicles, the receiver judges the trustworthiness of these messages; this process mostly depends on the sender's digital certificate revocation status. Each message that is sent by a vehicle carries the sender's digital certificate. The ability of each vehicle to check all Certificate Revocation Lists (CRL) to verify the sender's certificate in a timely manner not only is a high-level requirement but also a significant criterion in the process of revocation status validation. Hence, it is extremely desireable to develop an effective mechanism to accelerate the sender's authentication in VANETs.

Usually the authentication in a certificate–based system will be conducted by verifying the digital certificates that are received by an entity. In order to so, the receiver needs to check the revocation status of these certificates in the latest issued CRLs. This ensures that the certificate is not listed in any CRL and can therefore be considered a trustworthy entity.

Because the number of vehicles could be unpredictably large and each car could possess several certificates, the authentication process followed by the receiving entity could be quite overwhelming. In addition, due to the high number of certificates issued and revoked, as well as the fact that CRLs will keep all revoked certificates as well as new ones until they expire, CRLs could be large in size, rendering the search for

a revoked certificate in the CRL even more troublesome. For instance, some CRLs files that are issued by VeriSign, a company that operates a diverse array of network infrastructure and provides a range of security services, could reach sizes in megabytes, which is quite large for CRLs that usually reach tens of kilobytes [5]. This means that releasing, distributing and processing CRLs can become significantly time consuming and can incur high overheads in terms of communication and computation.

The situation is for the complication when a vehicle needs to verify multiple certificates that have been received from multiple senders simultaneously. Therefore, there is a need to design an efficient and powerful method that can overcome these issues. Using the same principles used in data mining (i.e., retrieving informative patterns from a large quantity of data), we integrated the clustering method used in data mining into the process of checking the certificate revocation status in CRLs. In order to achieve this, the well-known, k-Means clustering algorithm, which is a method that aims to partition n observations into k clusters, is employed.

1.2 Thesis Objective

The main objective of this thesis is to design and implement an authentication scheme for vehicular communication that is able to provide fast response time to peer vehicle authentication by integrating clustering techniques taken from data mining and applying then to certificate revocation status verification. Two goals are set before us: taking certificate's reputation into account and improving the clustering techniques we employed. Towards these objectives, the following relevant work is addressed in this thesis:

- A comprehensive revision of current authentication protocol will be given as well as certificate revocation status validation methods in Public Key Infrastructure (PKI). The related work to be studied provide an overview of the security implementation. Moreover, a detailed study of the classic k-Means clustering algorithm will be presented.

- An efficient protocol for retrieving the certificate revocation status will be proposed. The challenge of verifying a certificate with a quick response can be solved by adding two new attributes to future certificate structure standards.

- A comprehensive security analysis of the proposed protocol, with mathematical evidence will be undertaken. In this analysis, each authentication phases of the proposed protocol will be covered; i.e., phases of message signature and verification, CRL issuing, certificate re-signing and certificate revocation. A summary of the security analysis will be provided to illustrate the conclusion that highlights the strength of the proposed protocol.

- A performance evaluation will be conducted through a collection of simulations for both the proposed protocol and the lineal searching scheme, which is a typical means of retrieving the certificate revocation status: A comparison of these two validation schemes will be provided. The simulation will be performed in various scenarios and the results generated will be recorded and analyzed.

1.3 Contributions

The main contributions of this thesis can be broken down into three parts:

1. A new authentication scheme specialized in vehicular communication, Efficient K-Means Authentication (EKA), is proposed. The concept of clustering used in data mining is utilized in this scheme to guarantee the promptness of authentication. To our best knowledge, this is the first thesis that tackles the certificates revocation status problem by adopting data mining.

2. An initial centroids selection scheme in the k-Means clustering algorithm is proposed. Rather than deploying the original k-Means algorithm, we improve its initial centroids selection scheme to optimize the selection process.

3. The addition of two new attributes, credibility and issued date, is suggested for current CRL standard. As an extension to the existing X.509 certificate standard, credibility is added to the "Not Before" field, which was originally available in the digital certificates format in order to enable a CRL file to be cluster-able, thus accelerating the revocation status validation operation.

1.4 Thesis Organization

The rest of this thesis is organized as follows:

- Chapter 2 presents a background study of related works in the area of VANETs and its security issues, as well as data mining techniques, whose main goal is to discover meaningful knowledge from large quantities of data.
- Chapter 3 presents an over view of the proposed EKA authentication scheme. This starts with an explication of the new attribute "credibility", which consists of a detailed explanation of the reputation model and trust management in VANETs. Following this, the theory of basic k-Means clustering is provided, as well as an algorithm to help better understand its initial centroid selection. The concept of integrating clustering into the certificates validation as well as more details on the EKA are also provided.
- In Chapter 4, typical types of attacks against messages delivered in VANETs are discussed, followed by a set of security analyses that prove our scheme improves security against attacks.
- In Chapter 5, the effectiveness of our proposed technique is demonstrated via extensive simulation results. Details of the methodology used during the simulation as well as the analysis of the results obtained will also be provided.
- Chapter 6 shows the future research plan and draws the conclusion of the thesis.

Chapter 2

Background and Related Works

Various studies as well as an overviews of the current literature related to the subject of wireless networks, Mobile Ad hoc Networks, Vehicular Ad hoc Networks, security issues in VANETs, and data mining are discussed in this chapter.

2.1 Wireless Networks

Wireless technology is considered to be a revolutionary paradigm shift as it enables telecommunications between various entities without introducing the need for cables of any kind. In order to achieve this, in addition to the costly process of placing cables in equipment locations being avoided the introduction of exciting multimedia technology developments is made also available; This is especially notable in the communications and applications fields, such as smart phones and satellite communications. Aiming to provide ubiquitous, flexible, long distance and global communications in telecommunication areas, numerous different kinds of wireless network architectures have been proposed and implemented separately as solutions for online access. New applications such as the *Intelligent Transportation System* (ITS), disaster avoid systems, etc., have been introduced thanks to the development of wireless network techniques. For example, in the case of a disaster, a networks access maybe interrupted; but since the introduction of

simple and rapid reconstruction, the network's connection can be easily rebuild, even if the services of some communication links and nodes are unavailable at the time. In the following sections, some newly developed aspects of wireless networks will be discussed.

2.1.1 Wireless Technology

Wireless technology can enable communication via radio frequency communication, microwave communication or infrared short-range communication. Applications may involve point-to-point communication, broadcasting, cellular networks and other wireless networks. The electromagnetic spectrum can make use of light, colors, and electronic devices. As they are treated as a public resource, the frequencies of the radio spectrum that are available for communication are regulated by government organizations such as the the *Federal Communications Commission (FCC)* in the USA and *Industry Canada* in Canada. This can help to decide which frequency ranges can be used for what purpose and by whom.

It is believed that cognitive radios will be a better solution to increasing spectral efficiency in wireless networks. A comprehensive summary is provided in [6], aiming at addressing the fundamental capacity limits and related transmission technologies for varied wireless network paradigms. Side information usually consists of the knowledge of important information, such as activities, channels and messages from other nodes who shared the same spectrum.

With cognitive radio and spectrum access concepts having been introduced into the field, the spectrum sensing problem has attracted more attention. This problem is one of the most challenging aspects to the cognitive radio community. A survey of spectrum sensing protocols in cognitive radio is given in [7] by Yucek and Arslan. Most aspect of spectrum sensing problem is reviewed and multidimensional spectrum sensing concept is studied. In addition to these, challenges related to spectrum sens ing are explained and enabling spectrum sensing schemes are introduced, Finally, external sensing algorithms that associate to this problem are discussed.

Table 2.1: Comparison between IEEE 802.22 standard and other standards in IEEE 802 series.

IEEE Industry standard	Network type	Range	Maximum Data Rate	Frequency
802.22	RAN[1]	30km	31 Mb/s	54-862 MHz
802.16	MAN	1-2km	54 Mb/s	< 2.4 GHz
802.11b	LAN	33m	11 Mb/s	2.4 GHz
802.11a	LAN	20m	54 Mb/s	5 GHz
802.15	PAN[2]	10m	10 Mb/s	2.4 GHz
1. RAN: Reginal Area Network; 2. PAN: Personal Area Network.				

A high detailed overview of the under developed IEEE 802.22 standard for *cognitive Wireless Regional Area Networks* (WRANs) is presented in [8]. This standard is specified by the IEEE 802 Local Area Network (LAN)/Metropolitan Area Network (MAN) Committee. A comparison of IEEE 802.22 standard with other standards in IEEE 802 series is illustrated in Table 2.1. Aiming at provide broadband access in sparsely populated zones where wireless approaches cannot be economically served, IEEE 802.22 standard utilizing cognitive radio technologies to achieve this goal on the basis of the more commonly used VHF/UHF TV broadcast bands. This approach could reduce the large cost in both economic and social aspects and increase the usage efficiency of that spectrum.

2.1.2 Optical and Photonics Networks

Increasing demands for broadband services of Internet have triggered by its rapidly growth. For the further success of Internet, it is necessary to design a new way to provide broadband access. Optical networks are strongly able to provide high bandwidth as well as low latency. However, it is costly that using an optical network act as

broadband access network. At the same time, a certain number of nodes with also a fixed bandwidth in a pure wireless network cannot provide satisfactory QoS, because network throughput falls with the increase of quantity of users. In [9], Pan Li *et al.* propose a hybrid wireless network that integrate optical network and wireless network so as to provide the broadband access.

An interest of using photonics techniques's capability of broadband and low loss in the field of microwave signal's generation, distribution and control for applications has been germinated. Such applications include broadband wireless access networks, radar, satellite communications and so on. In [10], methodologies that have been proposed in the last few decades in the field of microwave photonics are studied, Yao puts a special emphasis on aspects of photonic generation, photonic true-time delay beam-forming and photonic analog-to-digital conversion etc. At the end, challenges and difficulties in practical application implementation and new research areas are also presented.

2.2 Wireless Ad hoc Networks

A Wireless Ad hoc Network is a decentralized type of wireless network. Ad hoc mode enables wireless devices to directly communicate with each other without a predefined central administrative infrastructure, (e.g., Wi-Fi access points). Instead, each node operates in a manner of relaying information for each other within their range, and therefore, the route to destination nodes will be dynamically chosen depending on the connectivity environment. Since, in most cases of ad hoc networks, such complete paths from origin nodes to the destination do not exist, traditional routing schemes would not work well in those scenarios because they need to establish a complete route before delivering data to the destination.

Increasing interest from both the scientific community and the industry were put on wireless ad hoc networks. At situations where fixed infrastructures are too difficult or even impossible to deploy, and also where the coverage of a Wireless Area Networks

(WAN), such as Wireless Personal Area Network (WPAN) and Wireless Local Area Network (WLAN), is required, WANET may find suitable application there. As mentioned before, unlike traditional wireless networks where the presence of stationary and reliable infrastructures such as base stations are needed to construct an end-to-end communication; ad hoc network is a technique that enables both portable and static terminals exchange data with each other. The assumptions of centralized management cannot be guaranteed: in best effort way, all involved nodes behave like terminal and router at the same time, and messages flows are transmitted through multi-hop routes between sources and destinations. In WANET, the cooperation among all the nodes is the key point behind such centralized administration absent wireless networks; and to achieve the objective of efficient resource sharing, self-organization and self-configuration at several protocol layers are required. In the following sections, some newly developing aspects of wireless ad hoc networks will be discussed.

2.2.1 Wireless Sensor Networks (WSNs)

An comprehensive study of applying WSNs to real-world dwelling monitoring is given in [11]. A series of design requirements are developed, which cover the design of the sensor network deployment and specify the capabilities for remote data access. Also a general architecture of the system is proposed to meet these requirement for habitat monitoring, and an instance using in seabird nesting environment and behavior monitoring is presented as well. In the future, the application of this architecture can serve for identifying important areas such as communications and health monitoring.

Playing a crucial role in surveillance area for automatic object detection, such as moving vehicles monitoring in traffic related fields, irrigation water monitoring within the agricultural industry, intrusion detection in security areas, and so on, WSN is a progressive research area with many academic workshops and conferences set up each year. It is inevitable that effective video compression and transmission of complex image or graphic data over the wireless network with high reliability in real time are compulsory

at surveillance systems. A WSN architecture based on line sensor framework that capable of capturing a continuous stream of one dimensional image is presented in [12]. Achieving faster processing with less storage and also considering bandwidth limitation when preserving the energy of each node, an associated one dimensional image processing algorithm is proposed as well.

2.2.2 Mobile Ad hoc Networks (MANETs)

For transmitting data in ad hoc networks, most researchers have argued the use of flooding-based routing mechanisms, but these mechanisms would consume energy and resources rapidly which could jeopardize a system's performance. To deal with such a dilemma, Spyropoulos *et al.* in [13] introduce a series of routing schemes that replicates data to be sent as several copies and route each copy independently towards a specific destination. Spyropoulos *et al.* also proved that not only were the proposed schemes a resultant in significantly less transmission and lower average delivery delays, but also maintained a high degree of scalability and good performance in large range grid scenarios.

It is critical to ensure the security in an ad hoc network, but also it is quite challenging due to the mobile nature of nodes and scarcity of a centralized administration. In order to solve this issue, Li and Liu proposed a distributed *ID-based Multiple secret Keys Management scheme* (*IMKM*) in [14]. This scheme operates through ID-based multiple secrets and threshold cryptography, without pre-shared certificate-based public keys. They proved that this scheme provides adequate efficiency for key update.

2.2.3 Wireless Mesh Networks (WMNs)

In wireless ad hoc networks, hierarchical mobility management schemes have been shown sufficient effects when providing low latency hand off. A investigation about the suitability of hierarchical mobility management schemes in WMN was performed in [15],

the results demonstrates that those scheme are not directly suitable due to the reason of WMNs have graph based topologies instead of tree based topologies. In this paper, a scheme for locating the optimized deployment of *Mobility Anchor Points (MAP)* in WMNs is presented. With experimental results, it is argued that Hierarchical Mobile IP (HMIP) is applicable on WMNs with optimized deployment of MAPs, and also able to reduce the optimal packet loss rate and hand off latency.

WMNs have been implemented in many urban areas where a large number of business units are located because of the simplicity in deployment and low cost. Establishing secure communications among different entities, e.g., mesh routers and mobile clients is chanllenging. The situation becomes even more difficult when roaming to other mesh networks. In [16], a new secure protocol with ID-based cryptography that aims at securing communications in large-scale multi-domain WMNs. In this protocol, mobile clients can get access to secure communication easily and conveniently even when roaming.

2.3 Vehicular Ad hoc Networks (VANETs)

Vehicular ad hoc network is an emerging type of network that allows moving vehicles on the roads to act as wireless nodes within a wireless ad hoc network. A typical example of vehicular communication systems is shown in Figure 2.1 . Compared to traditional ad hoc networks, VANETs consider vehicles and other entities on the road as the communicating nodes in their system, regardless whether they act as terminals or relays. To connect to and establish communication between entities, the effective communication range is up to about 300 meters from one vehicle to another. As some vehicles can, at any time, move out of this communication range with one another while moving, others may join in once they enter the effective communication range. With the demands of improving road safety and the development of electronic components, vehicle manufacturers have equipped their products with dozens of interconnected on-board processors and accessories (e.g., GPS, sensors, etc.), all of which turn vehicles into what is known as “computers on wheels”.

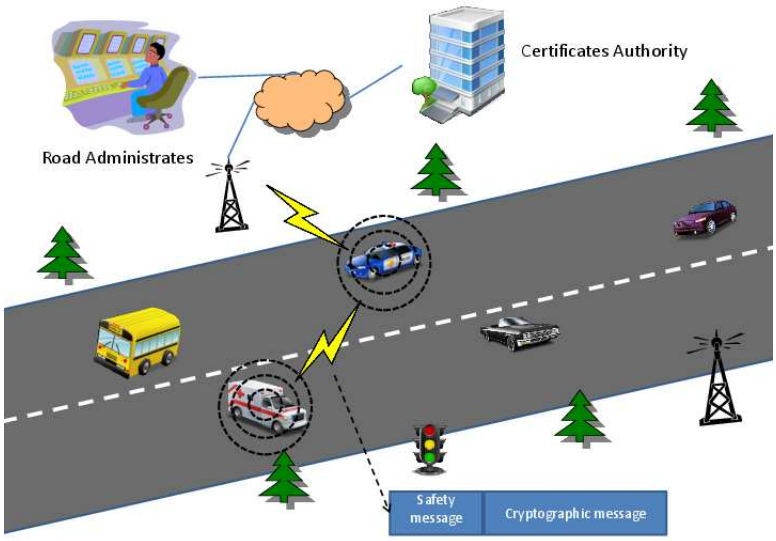

Figure 2.1: Overview of the structure of VANETs.

One of the key communication patterns in VANETs is beaconing, which is the process of periodically broadcasting status information. Thus, vehicles on the road that have communication devices equipped can use this process to gain information about other communicating entities. Usually, information that is contained in the beacons' messages include a vehicle's unique identifier, geographical location, velocity, etc. Obviously the process of beaconing is helpful to vehicles since it provides vehicles with an awareness of the surrounding environments. This awareness must be able to guarantee two aspects: the correct assessment of safety related circumstances and precision of congestion levels on road for traffic safety and efficiency purposes. When considering VANET applications for automated driving, these two aspects could be utilized as a key parameters as well. Moreover, beaconing is usually considered as mission critical as driving decisions may be made based on a beacon message's information.

The other key communication pattern is routing. In VANETS, searching for routing routes is challenging due to the dynamic nature of moving nodes. Numerous different routing protocols have recently been proposed for routing in VANETs, which assume a pure ad hoc architecture. Since VANETs and MANETs share the same routing princi-

ples and have many similarities, such as communication that does not rely on a fixed infrastructure, self-organization and self-administration, most ad hoc routing protocols are applicable in VANETs, such as *Ad-hoc On-demand Distance Vector* (AODV) [17] and *Dynamic Source Routing* (DSR) [18], which are principally designed for MANETs use. These routing protocols do not maintain routes unless message forwarding is needed. Hence, in scenarios with a small quantity of communication flow, these routing protocols can significantly reduce overhead. However, VANET differs from MANET because of its highly dynamic topology. In the past few years, several studies and simulations have been performed to compare the performance of these routing protocols in different traffic situations [19][20][21][22][23][24]; These simulation results demonstrated that most ad hoc routing protocols (e.g., AODV and DSR) are effectively not applicable for VANETs because of high node mobility in VANETs.

2.3.1 Standards Developments of VANETs

In order to regulate the *Wireless Access in Vehicular Environments* (WAVE), the IEEE developed a series of standards, namely the IEEE 1609 Family of Standards. These sets of standards were built upon the IEEE 802.11p, which is an improvement of 802.11 for implementing WLAN computer communication. In addition to outlining an administrative model and communications architecture, it also defined the security mechanisms employed as well as physical access. According to the conclusion from Grafling in [25], the IEEE 1609 Family of Standards can maintain stable performance even when faced with a high density of vehicles, which provides a platform for a variety of applications to be built upon.

Dedicated Short-Range Communication (DSRC) is another series of protocols and standards that is specifically designed for Vehicle-to-Vehicle (V2V) and Vehicle-to-Roadside (V2R) wireless communication within a certain range. DSRC has the potential to host a variety of new applications in vehicular environments, the most critical of which are safety-related applications; these can in clude collision avoidance, speeding warnings,

etc. In [26], Kenney believed that DSRC is meant to act as a supplement for cellular communications in vehicular environments by offering high data transfer rates and reducing transmission latency. Furthermore, DSRC is able to separate comparatively small communication zones during vehicular communication.

2.3.2 Security Issues of VANETs

Security issues are always a focal field in VANETs. In addition to setting up communication standards, automotive industries have also spent a tremendous amount of effort on the development of traffic condition related systems, and more specifically in ensuring reliable message delivery between involved entities. However, even so, malicious nodes may still compromise the communication operation of vehicles and can potentially steal information, which raises the issue of privacy. In [27], Zhang discusses the challenges of trust management in VANETs due to their decentralization. However, a number of current reputation models that exist in other ad hoc networks have been studied and a suggestion for a powerful trust management scheme has been proposed. One such model has been suggested by Almulla *et al.* in [28].

It is undisputed that vehicular communication systems have the most potential for high profit in the domain of wireless ad hoc networks considering the tremendous number of vehicles on the road and the large potential profit gained from them. The incorporation of on-board units into vehicular communication systems would provide substantial business opportunities. However, vehicular communication systems also present challenges to academic researchers with regards to security. Raya and Hubaux in [29] conducted a detailed analysis of securing vehicular communication and also depicted a framework for an ideal vehicular communication system architecture. In addition, a review of their proposed framework as a robust system is also provided.

In VANETs, it is important to develope a scheme to determine if the quantity of nodes that are reporting an event is larger than a threshold or not. Many applications reach an agreement among the vehicles depending upon a threshold number of warning

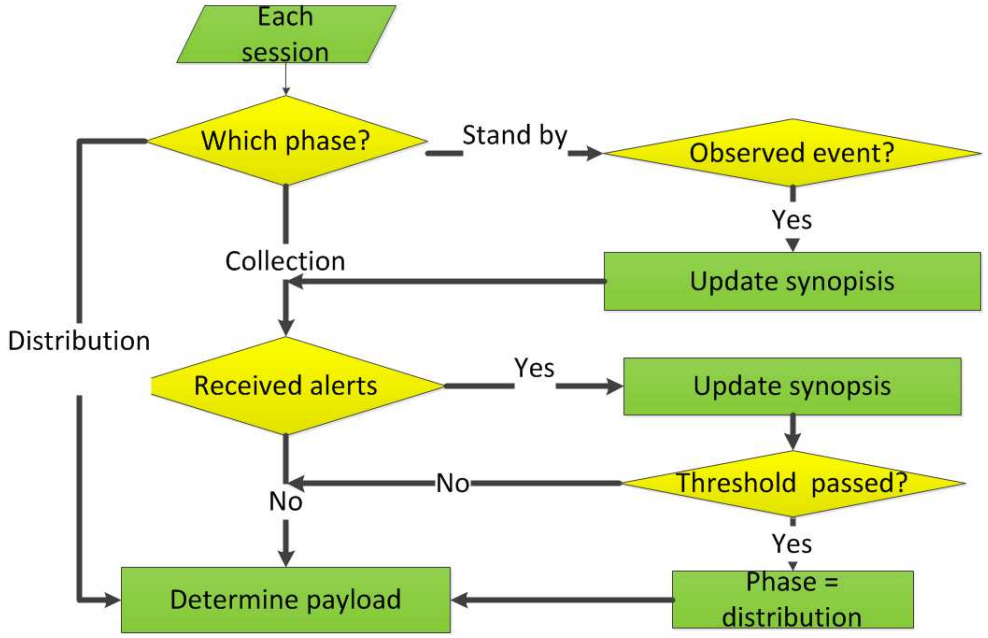

Figure 2.2: Phase transitions in threshold-based applications.

reports, thus they are able to regulate the validity of an event or avoid the abuse of emergency beacon broadcasting. An efficient threshold-based event validation protocol is presented in [30]. Analysis and simulation results prove that this protocol is fairly accurate despite the presence of attacks and is also capable of collecting and distributing warning notifications in less than one second with negligible requirements on network bandwidth. An illustration of this proposed threshold-based event validation protocol is illustrated in Figure 2.2.

One challenge related to security issues in VANETs is the ability to detect the anonymous insider who conducts malicious acts. This challenge presents itself because of the difficulty of the conventional authentication mechanisms deployment. Most previous works concentrating on this challenge do not take privacy issues into account. In [31], two new protocols that aim efficiently at detecting and excluding the anonymous misbehaving insiders within VANETs are proposed. These protocols have proven, through extensive simulations, that they are capable of achieving rapid reaction and high accu-

racy, for both the eviction and the revocation of malicious vehicles, all at an acceptable cost.

2.3.3 Routing Protocols of VANETs

In [19], the dilemma of routing in VANETs and the most recent research addition to this is discussed and surveyed. It is believed that optimizing vehicular communication methodologies and appropriatly using the available wireless bandwidth will lead to the success of VANETs. A fundamental and practical investigation aimed at discovering the possibility of reducing the overload on the wireless channel, is addressed by Mittag *et al.* in [32]. Rather than transmitting periodic beacon messages through one hop with high transmission power, this would attempt to transmit messages over multiple hops with less transmission power. In particular, Mittag *et al.* investigate the bandwidth saving method by integrating previously forwarded beacon messages into the next transmission. To evaluate the effects of packet collisions and channel fading, a stimulative comparison of one hope and multi-hop beaconing is performed. The results show that the assumed savings are difficult to achieve under practical wireless channel conditions.

The networks in VANETs can exhibit variant behavior at different times, i.e., either fully connected or sparsely connected. In [20], for the purpose of investigating the connection loss phenomenon and its network characteristics, Wisitpongphan *et al.* propose a comprehensive framework with actual empirical traffic data. Additionally, Wisitpongphan *et al.* also demonstrate in the case of VANETs safety applications, the traditional ad hoc routing protocols such as DSR and AODV cannot work well with such long re-healing times. Therefore, a new ad hoc routing protocol is needed.

2.4 Authentication

Many security mechanisms for wireless networks have been introduced in the literature. Authentication is considered as the first defense against attackers from compromised

sources and plays a important role in the entire wireless networks system. An authentication scheme will exchange a series of messages that carry verified and identified information of entities. The exchange is accomplished through two methods of cryptography: Symmetric or Asymmetric. Not only is it considered as the first line of defense against attacks, authentication also builds a basis for accomplishing privacy and confidentiality. Ad hoc networks differ from other wireless networks because of their dynamic topology, thus, efficient and secure authentication schemes are essential for VANETs in order to guarantee QoS. A sample illustration of the certificate-based mutual authentication from [33] is shown in Figure 2.3.

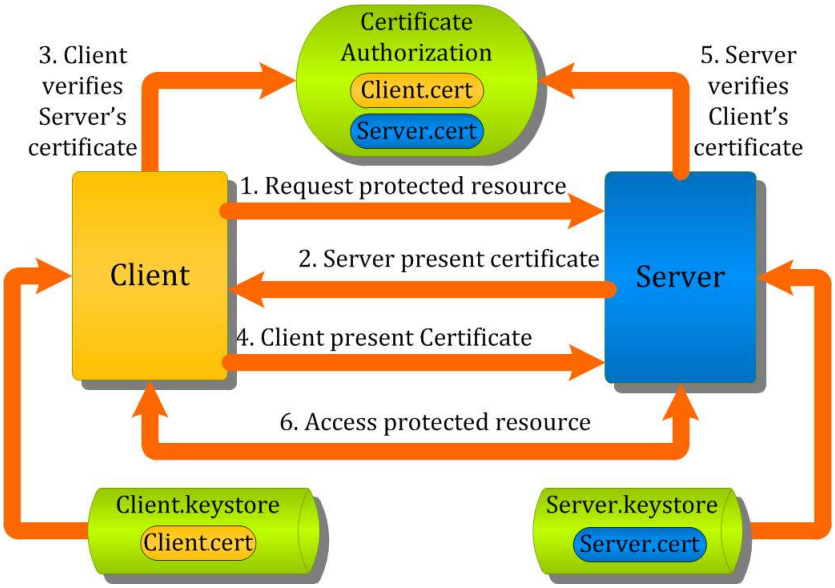

Figure 2.3: Certificate-based mutual authentication.

Safdar *et al.* in [34] presented a novel authority authentication protocol for VANETs, entitled *"Randomly Shifted Certification Authority Authentication protocoL (*RASCAAL*)"*. RASCAAL takes Medium Access Control (MAC) characteristics into account and takes advantage of the *Carrier Sense Multiple Access with Collision Avoidance (*CSMA/CA*)* medium access rules in its operation. RASCAAL forms dynamic clusters without using pre-defined cluster heads, as they can increase the network's vulnerability.

Considering the increased number of mobile users pursuing universal access to a wireless network, the inter-domain migration is also a challenge to authentication. In a situation where the CA is absent, an authentication mechanism is necessary between newly-joined mobile users and inexperienced domains. Chang and Tsai in [35] suggested a novel self-verified authentication scheme that utilizes *Elliptic Curve Cryptography (*ECC*)* in order to achieve security goals. Although the computational complexity of ECC leads to lower performance efficiency, the scheme provides a degree of security to its users that can be considered as a trade-off to its decreased performance.

Recently, researchers' efforts have been focused on using physical layer information to improve wireless security. It is believed that the integration of security with the physical layer characteristics could be a potential complement to enhancing wireless networks security. Zeng *et al.* in [36] conducted a survey of several non-cryptographic methods that use physical layer information in both static and mobile wireless networks. In addition, they also discuss the advantages as well as limitations and challenges of its implementation. Wen *et al.* in [37] introduced a message authentication framework that uses physical layer data to perform cryptography. It is an integration of conventional authentication methods and physical layer authentication schemes by means of temporal and spatial information in physical layer channels to achieve faster authentication.

In addition to those mentioned above, several additional authentication mechanism have been introduced. A novel authentication mechanism tailored for ad hoc network environments was explored in [38], titled "*nested one-time secret mechanism*". Used in mobile Worldwide Interoperability for Microwave Access (WiMAX) authentication, the IEEE 802.1X process incurs a long delay in hand off. To eliminate the unnecessary authentication cost, a key caching mechanism is proposed in [39]. A simple and convenient authentication scheme to be utilized in the global mobility network (GLOMONET) is presented in [40]. An innovative authentication protocol that is aimed at reducing authentication cost is introduced in [41].

2.4.1 Digital Certificate

A digital certificate, also known as a public key certificate, contains identity information encrypted by a public key as well as a digital signature from a trusted CA that ensures the legitimacy of this public key.

One of the most widely used standards for digital certificates is the X.509 standard, which is a well-known and prestigious standard. It was developed by The International Telecommunication Union Telecommunication Standardization Sector (ITU-T) for both Public Key Infrastructures (PKI) and Privilege Management Infrastructure (PMI). It specifies both contents and structures of public key certificates, Certificate Revocation List, and certification path validation algorithms. PKI puts more emphasis on authentication: only in this way will a user be able to prove that he has a private key that matches the public key and can therefore be authenticated. The latest version, X.509 Version 3 certificate, was released in May 2008. The well-known X.509 standard specifies all necessary information that consist of an X.509 certificate and defines its data pattern. In addition to a digital signature from a CA, X.509 certificates are composed of specific contents, listed below in Table 2.2.

Harn and Ren in [43] introduce the concept of *Generalized Digital Certificate (*GDC*).* These types of certificates do not have a public key from the user and this improvement makes the key management in GDC-based schemes easier to maintain. Harn and Ren also proposed a discrete logarithmic-based and integer factoring protocol that can help in the authentication process.

Holohan and Schukat in [44] present a PKI-based mechanism for ad hoc networks. This mechanism limits the amount of initial trust information that can be stored on devices, especially for those resource-constrained devices such as smartphones. Moreover, the benefits also include compatibility, a smooth upgrade from existing protocol stacks such as those standards that have been specified by IEEE 802.15 and ZigBee.

Roy-Chowdhury and Baras in [45] address a lightweight symmetric key certificate scheme termed “TESLA”. In this scheme, messages are always authenticated with a

Table 2.2: Structure of an X.509 Certificate.

Certificate certificate fields,

- **Version** explicit version default v1,
- **Serial Number** certificate serial number,
- **Algorithm ID** algorithm identifier,
- **Issuer** name,
- **Validity** validity,
  - **Public Key Algorithm** public key algorithm,
  - **Subject Public Key** subject public key,
- **Subject** the entity identified,
- **Subject Public Key Info** subject public key info,
  - **Not Before** the date when certificate first valid from,
  - **Not After** the certificate expiration date,
- **Issuer Unique Identifier (optional)** implicit unique identifier,
- **Subject Unique Identifier (optional)** implicit unique identifier,
- **Extensions (optional)** explicit extensions,
  - **...** reserved for future use,

Certificate Signature Algorithm algorithm identifier,

Certificate Signature signature value.

(From RFC 3280, *The Internet Engineering Task Force (IETF)* [42])

sender's MAC information prior to delivery. With the use of symmetric MAC functions and hardware processing power, this scheme is considered lightweight and efficient. Roy-Chowdhury and Baras also compare its performance to other public key-based digital certificate schemes to prove its efficiency.

2.4.2 Certificate Authorities

CAs are essential in order to implement PKI solutions for the purpose of improving. The CA's main responsibilities related to public key certificates are as follows:

1. *certificate generation*
2. *certificate distribution*
3. *certificate renewal*
4. *certificate revocation*

Several critical responsibilities are overseen by the CA. However, for certificate revocation in VANETs, the revocation scheme can be deployed in two methods: centralized and decentralized. The former indicates that a central authority is the only entity responsible for making a revocation decision, whereas in the decentralized method, the revocation decision could be done by multiple vehicles in order to detect neighboring vehicles behaving in a malicious manner. In this chapter, primary focus is on those revocation schemes that operate in a centralized manner. In other words, where the CA is the only authority that is able to determine if a certificate needs to be revoked.

2.4.3 Revocation of Digital Certificates

In cryptography, the CA issues certificates to trusted users. However, due to some entities potentially having malicious behavior, these entities lose a CA's trust and are noted as untrustworthy entities. Once an entity's trust is lost, the CA revokes its certificates and

adds this entity in a Certificate Revocation List. CRLs are lists of revoked certificates that are not to be relied upon. These CRLs are released by the CA for distribution in order to inform other entities and applications not to trust holders of these certificates. It is obvious that the CRLs should be distributed in a manner both quick and secure.

Haas *et al.* in [46] present a way by which to reduce the CRL size and provide a computationally efficient method to validate whether a certificate is listed in a CRL or not. In addition, a lightweight CRL distribution mechanism was introduced. The validation of certificates is more complicated in ad hoc networks due to constant link disconnections between entities and the CA. For this reason, Forne *et al.* in [47] provide a new solution to overcome the shortage of infrastructure and resource scarcity in mobile devices. A survey of different certificate validation mechanisms in MANETs is discussed as well. Nowathkowski and Owen in [48] address scalable methods for distributing CRLs as they relate to multichannel operations in the IEEE 1609.4 standards series. From the simulation results, it was proven that these methods work very well regardless of varying vehicle densities.

2.4.4 Certificate Revocation Lists

CRLs carry important information regarding revoked certificates in PKIs. In order to make sure that the certificate is valid, it is necessary to frequently check CRLs . In wired environments, such CRLs are usually stored in a centralized CA so that they can be easily accessed by users. But in a wireless network with the instability of communication links and high mobility of nodes such as MANET, numerous CAs have to be used. Moreover, there is a possibility that sometimes users may not have access to data that is stored in a particular CA because of the user's mobility; therefore the distribution of the latest version of the CRL becomes important.

As we mentioned, CRLs are issued to all nodes in networks in order to maintain the list of revoked certificates. Certificates may be revoked [49] due to the following reasons:

1. *Key Compromising*: The private key of either the user or the issuer has been compromised or holds suspicious characteristics.

2. *Change of Affiliation*: The relationship between an entity holding a certificate and the issuing CA does no longer exist.

3. *Termination of Operation*: The certification is no longer needed because the purpose originally assigned by this certificate was changed or completed.

Currently, each CRL file carries information that is displayed in two tabs: the General tab and the Revocation List tab. The General tab shows information about the CRL itself, such as the issuer, the effective date of this CRL, the time when the next CRL will be released, and so on; The Revocation List tab, on the other hand, displays the serial numbers of certificates that have been revoked and the date when they were revoked, both of which consist of the most important contents in a CRL. Two examples of a General tab and Revocation List tab in existing current standards is illustrated in Table 2.3 and 2.4.

Table 2.3: An example of a general tab in the existing PKI standard.

Field	**Value**
Version	V2
Issuer	VeriSign Class 3 Secure Server CA -G3
Effective date	April-21-12 5:00:31 AM
Next update	May-05-12 5:00:31 AM
Signature algorithm	sha1RSA
Signature hash algorithm	sha1
Authority Key Identifier	KeyID = 0d 44 5c 16 ...
CRL Number	05 89

Table 2.4: An example of a revocation list tab in the existing PKI standard.

Serial number	**Revocation date**
12 fa d6 99 bb 6a 76 4f e3 a8 21 d0 9b 51 7a 5d	April-01-12 1:04:22 PM

2.5 Data Mining

As a relatively recent and interdisciplinary field of artificial intelligence, data mining aims at digging through and discovering meaningful knowledge from large quantities of data. The applications of data mining have risen dramatically over the last decade, and recently have attracted tough competition when it comes to the quality of information, since information collection plays a critical role during the decision making process. There are two main concerns in the data mining process: what has been provided and the speed of its delivery. Not only has the research community shown growing interests in this area but also governments and industries have shown an interest in order to serve different purposes.

In [50], Liu and Yu discuss feature selection concepts and complete a survey of existing feature selection algorithms from the perspective of clustering. In addition, guidelines on how to use featured selection algorithms has been provided, and a unified categorizing framework and platform has been proposed to assist in selecting the appropriate algorithm. Ultimately, Liu and Yu's intent was to build an integrated system that simplifies the intelligent feature selection.

Cai *et al.* in [51] introduce a clustering method in order to group documents with similar attributes. Since some documents may have high dimensionality, these documents can be mapped into a lower-dimensional semantic space by means of *Locality Preserving Indexing (*LPI*)*. Differing from conventional document clustering algorithms, this algorithm focuses on exploring the information of geometric and discriminating structures as the basis of document clustering.

Rather than providing a specific algorithm, Cao *et al.* in [52] developed an effective and general approach to mining informative patterns from several relevant large data sources that consist of multiple aspects of information.

Retrieving informative patterns in real-world large distributed systems could be comparatively costly because of the extremely large scale of some systems as well as the high cost of communication. Wolff *et al.* in [53] propose a two-step approach for clustering in large distributed systems in order to tackle this bottleneck. This approach adopts an efficient monitoring algorithm that monitors the cost.

Chapter 3

Efficient *K*-Means Authentication

In this chapter, we propose a fast certificate revocation status validation scheme for authentication in VANETs. We call this scheme the efficient k-Means authentication. This k-Mean clustering-based scheme can accelerate the authentication process with greater efficiency, with the acceleration achieved by adopting the following two aspects: the introduction of new elements in CRLs and by adopting the k-Means clustering algorithm with enhanced centroids selection. In virtue of this accelerated process, successful validation can be achieved.

3.1 The Foundation for Credibility: Reputation

In the first section, we will discuss the prerequisites for the credibility of the k-Means scheme. We use the term "credibility" to represent the reputation value of a node. Credibility originally refers to the degree to which a nodes is believed or trusted as a source of message. Specifically, a comprehensive study of the existing reputation models in VANETs is presented, with their key issues briefly summarize as well. In VANETs, a critical issue for both vehicles and RSUs is how to decide to trust the specific vehicles and messages based on evaluating the behavior of the vehicles. Academics proposed a new technique that aims to avoid compromising the system by malicious acts is gaining

popularity, and is known as the reputation system, or the reputation model. In this section will elaborate on reputation systems, followed by an introduction to two new attributes in the Revocation List tab proposed for future CRL standards.

3.1.1 The Definition of the Reputation System

The goal of the reputation system is to measure the role of nodes in wireless networks, and to perceive the risk of potential malicious behavior. The system then generates a reputation value that characterizes the trustworthiness of each node: Following this, it uniformly circulates and distributes the available reputation value of a node over the network as long as it is available. Based on these values, the nodes can take advantage of the reputation value to make a decision on whether or not a peer node is reliable and trustworthy when they must choose a set of trustworthy nodes for the purpose of communication. This essentially discourages untrustworthy behavior. Because both imply the similar concepts, trust and reputation can be used interchangeably. Various definitions of reputation are featured in the literature. For instance, in [54] Liu and Issarny addressed an ideal reputation system for MANETs, which they claim should have the following properties:

Valid: Users are able to distinguish effectively trusted from untrusted users through the reputation system.

Distributed: The assumption of access to any trustworthy entity or centralized storage for reputation systems is not guaranteed.

Robust: The system is capable of resisting attacks.

Timely: The information of the system should be live and exhibit the latest trustworthiness of an entity.

Resource-saving: The limited computation capability and storage space of some terminals in MANETs should be considered.

Because any infrastructure for data transmission is not guaranteed in MANETs, their communication, such as packet forwarding, is accomplished through the cooperation amongst neighboring nodes. This communication mostly depends upon the reputation value of these neighboring nodes. Therefore, the assessment of the reputation value of a node is of utmost important. A way to summarize the reputation of a node is by providing information on the node's past participation as a basis for evaluating its potential future behavior as well as observing its involvement in other communications.

3.1.2 Reputation Model

Reputation plays an important role in MANETs. Works on reputation systems generation are divided into three main categories :

1. *Evidence Collection:* collecting a node's past participation performance as feedback for reputation values generation.

2. *Reputation Formation:* apply a mathematical algorithm to calculate a score, i.e., reputation value of the node with the collection of involvement history as input.

3. *Decision Making:* nodes decide to trust a target node based on a threshold value.

In the first phase, a set of nodes specify a target node and gather sufficient statistics of past interactions in which it was involved. The recorded behavior patterns of a node is gathered either through direct observation or indirect assessment from other nodes. The main concern during this phase is the representation of a node's past performance. In the second phase, nodes apply an existing algorithm to calculate the reputation value with the collected evidence gathered in the first phase as input. Finally, the nodes decide to trust the subject node or not. A threshold is required to make this decision. If the reputation value generated in the second phase is larger than the threshold, the subject node is trustworthy; otherwise, if it is below the threshold it is not considered trustworthy.

3.1.3 Reputation Models in MANETs and VANETs

VANETs, as one important application of MANETs, shares some similar properties with MANETs, e.g., decentralization, mobility and openness. But differences still exist between them. The quantity of nodes in VANETs are often much larger and can potentially reach a level of millions of vehicles. In addition, the network communication overhead may be high in heavy traffic environments. Moreover, since vehicles move fast, the topology of VANETs usually changes rapidly.

In VANETs, trust management is more challenging when compared to MANETs. More specifically, one fundamental assumption in many existing MANETs reputation models in MANETs is that reputation values are always accessible before the route of package forwarding is founded. Practically, however, unless a reliable route is confirmed, trust cannot be formed, maintained or retrieved. This is also one of the reasons that trust management is more difficult to establish in a topology with a constantly changing environment such as VANETs.

Most of the reputation models in MANETs focus on sparsely distributed nodes which models the reputation value of nodes by collecting previous trust evidence about them from other nodes with which they have communicated as well as through some midway nodes. This is certainly difficult to achieve in VANETs since the environment of VANETs is often very large and, considering the limited time given for decision making, searching for necessary trust evidence may turn into an impossible task. Additionally, nodes in MANETs may not cooperate sometimes. In resource limited environments, nodes can refuse to cooperate so as to save energy or for other selfish reasons. Finally, trust in MANETs is not automatically transitive; for example, a node named “Alice” trusts a node named “Bob” and “Bob” trusts a node named “Tom”. This does not indicate that “Alice” trusts “Tom”.

As opposed to those in MANETs, the objectives of trust management in VANETs are not limited to reliable message delivery. Two main goals of VANETs are to increase the safety of driving and reduce the congestion of traffic, by sharing information about traffic

conditions amongst peers. Aiming to help peers detect false information and make correct driving decisions, trust management presents more of a challenge than reliable message delivery. Many characteristics of a reported event have to be taken into account, such as the time, the location, and the critical levels. Thus endeavoring to directly apply the reputation modeling that is proposed for MANETs to VANETs becomes impracticable and worthless.

Only few reputation models have been recently proposed for implementing trustworthy data sharing in VANETs. In this section, these are summarized and their issues are elaborated on. It is noteworthy that, a substantial amount of effort, e.g. [55][56], has been spent in ensuring security and the preservation of privacy in trust establishment in VANETs. Most of proposed systems rely on security infrastructures and that make use of digital certificates. We will focus on reputation models that are not completely built upon the static security infrastructures and that can therefore be more portable deployed. In those models, trust relationships are established mostly based on past communication patterns. Information about a traffic environment sent by other nodes may also be collected to determine the exactness of the messages. These models consist of three main classes: entity-oriented reputation models, data-oriented reputation models, and combined reputation models. They are described as follows:

1. *Entity-oriented reputation models*: reputation models of this class mainly concentrate on modeling the trustworthiness of peer nodes.[57]

2. *Data-oriented reputation models*: unlike entity-oriented reputation models, reputation models of this type focus on appraising the trustworthiness of data. Normally, long-term relationships between peers do not exist in these models.[58][59]

3. *Combined reputation models*: combined reputation models not only estimate the trustworthiness of data through peer trust, but also maintain peer trust for a certain period of time. [60][61]

3.1.4 Calculation of Credibility

As for the calculation of the credibility, a method proposed by Ren and Boukerche in [62], termed the *Generalized Reputation Evaluation (*GRE*),* is employed. As mentioned before, the reputation value of a node is usually calculated in most reputation models by applying a mathematical algorithm, which commonly considers a single influencing factor and uses a linear function. However, the novel trust-based reputation evaluation methods, GRE, is able to update the reputation values by taking two factors of a node into account: its residential time T and the recent activity ra. Before elaborating upon the GRE, the preliminary assumptions and concepts will be introduced first.

- The characteristic of bi-directionality exists in all links between two wireless nodes in VANETs.
- Sufficient computational power to go through all operational steps is always guaranteed for every node in VANETs.
- A trusted CA outside of the VANETs is responsible for issuing and updating the public keys and private keys to vehicles and RSUs as well as receiving reputation value reports that are sent by nodes.
- The concept of *community*: a community indicates a central node and all of its one-hop neighboring nodes; some malicious nodes might exist within in a community.

A node's previous historical trust records have an important effect on its current computation of reputation value, therefore the recent trust of the node n_i is represented as rt in order to reflect the past behavior of n_i. Two influencing factors are considered: the residential time T and the recent activity ra. When a node n_i stays in an inexperienced community, the residential time of the node shows the degree of its trustworthiness. The longer n_i stays in the community, the longer T_i is and thus the more trustworthy n_i is.

Once a malicious node is detected, it will be isolated from the community and therefore cannot be maintained. More precisely, the time T of all nodes is measured in a time unit such as millisecond. The amount of the node's recent activity is recorded, with the reputation depending on the period of time that node has remained in the community as well as on the past trust of this node. After that, a mathematical equation is defined

$$\begin{cases} 0 < rt < 0.5 & & 1 < \alpha < 2 \\ rt = 0.5 & \mu = \alpha \times rt & \alpha \approx 2 \\ 0.5 < rt < 1 & & 2 < \alpha < 3 \end{cases} \tag{3.1}$$

Where α is a decimal and its value determined based on rt as well as the individual node. The reason for using rt is to yield a value close to 1 for nodes that have a moderate trust ($rt = 0.5$), a value less than 1 for those with a lower trust ($rt < 0.5$), and a value larger than 1 for those that have a higher trust ($rt > 0.5$). We then use ε stand for the time factor

$$\varepsilon = \kappa^T \times ra \tag{3.2}$$

where κ is a discount factor that is above 0 but below 1, and when ra denotes the node's recent events, which can be a successful package dispatching or an intended exaggeration. As a final point, the *Trust* metric, i.e, reputation value, is assessed as follows:

$$Trust = \begin{cases} \lambda \times \frac{1-\mu^{1+\varepsilon}}{1-\mu} & \mu < 1 \\ \lambda \times \frac{\mu^{1+\varepsilon}-1}{\mu-1} & \mu > 1 \end{cases} \tag{3.3}$$

Where λ is a scaling factor used to keep the reputation value *Trust* at a value between 0 and 100. The values for $\varkappa$ and λ will be carefully chosen individually by each node. Consequently, the increase in reputation value will have three shapes base on the previous reputation value and the period that the node has stayed in the community. If node n_i has had good historical trust performance, then its current reputation value will rise

rapidly. However, if n_i has fewer trust records, its reputation value will growth at a slow pace. Finally, for a node n_i that has a intermediate trust credits, its trust will also increase at a moderate pace. After the value of *Trust* is determined, nodes will report this new value of n_i to CA via vehicular communications.

Another issue is in maintaining the community. A method similar to that used by the Ad hoc On-Demand Distance Vector *(AODV* [17]) is utilized for the maintenance of the community in that the central node periodically broadcasts a HELLO message. At the end of each session, the central node will clear the value of T and ra respectively and use each node's current reputation value *Trust* to substitute its corresponding recent trust rt. Through introducing this efficient reputation value calculation model, which is neither difficult nor intricate, the proposed certificate revocation validation scheme Efficient K-Means Authentication *(*EKA), to be presented in following section, can be applicable for VANETs.

3.2 New Attributes: Credibility and Issued Date

By including a credibility factor to the existing Revocation List tab, certificates can be categorized into different levels of trustworthiness, providing a basis for the acceptance or rejection of these certificates.

The numerical ratings for certificate credibility are set based on the *Trust* reports corresponding to the last CRL issued and have number designations with a rating scale from 0 to 100, as show in Table 3.1 . The CA collects those reports and calculates the mathematical average of *Trust,* then assign it as the credibility of the certificate that node n_i possesses. Any certificate with a ranking lower than "Good" will be considered as suspicious or malicious. In the illustration of CRL content in Table 3.2 , credibility is listed as the third attribute in the Revocation List tab, after Serial number and Revocation date.

In addition to credibility, the "Issued Date", which is a modified version of the "Not

Table 3.1: Credibility rating value.

Numeral Ranking	Meaning
A (e.g., 100 ~ 80)	Excellent
B (e.g., 79 ~ 60)	Good
C (e.g., 59 ~ 50)	Suspicious
D (e.g., 49 ~ 0)	Malicious

Table 3.2: An example of a Revocation List tab for future standards

Serial number	Revocation date	Credibility	Issued date
12 fa d6 99 bb 6a 76 4f e3 a8 21 d0 9b 51 7a 5d	April-01-12 1:04:22 PM	53	February-07-12 7:00:00 PM

Before" field in X.509, is proposed. The difference between these two attributes lies in the fact that the "Not Before" attribute does not take into consideration the date the certificate was issued but only the effective start date and time of the certificate. For the "Issued Date" attribute, the value assigned depicts the date and time a certificate was issued, regardless of its starting time and date. In this work, the Issued Date attribute is added to the Revocation List tab as the fourth attribute, shown in the last column of Table 3.2.

3.3 *k*-Means Clustering

In order to ensure a smooth transition into our proposed centroids selection approach, there is a need to understand how the original k-Means clustering method works.

The reason that why *k*-Means clustering algorithm is chosen among other clustering

algorithm is because of its unsupervised learning. it is a classical partitioning algorithm, for clustering n data points into k discrete clusters C, with the cluster C_j containing n_j data points. Each cluster has a centroid, which represents a central vector used to assign different entities to that specific cluster. k-Means picks an initial centroid randomly and then uses Eq. (3.4) to determine the next cluster centroids

$$L = \sum_{j=1}^{k} \sum_{i=1}^{n} \|x_i - \mu_j\|^2 \tag{3.4}$$

Where x_i is a vector denoting the x_i-th data point, μ_j is the centroid of data points in C_j and L is the distance for each data points to all centroids.

The original k-Means clustering algorithm is described in Algorithm 3.1:

Algorithm 3.1 Original K-Means Clustering Algorithm

Require: Input the number k of cluster centroids.

Ensure: Output k cluster .

1: Get $k = numberofclusters$

2: Get $X = (x_1, x_2, \ldots, x_3), x_i \in R^d$

3: **for** $j = 1$ to k **do**

4: select $\mu_1, \mu_2, \ldots, \mu_k$ randomly

5: **end for**

6: **for** $i = 1$ to n **do**

7: determine $\mu_j = \{\mu_j | argmax \sum_{j=1}^{k} \parallel x_i - \mu_j \parallel^2\}$

8: **end for**

9: Assign x_i to μ_j

10: After all data points have been assigned, recalculate the position of the centroids.

11: Repeat step 4 to 10 until all centroids are convergent

The centroids are considered as converged if their positions do not change any more after a number of iterations. According to [63], the algorithm can be stopped once the t -th iteration has been achieved, with an initial given threshold of ε and if those positions

have been validated by the following inequality Eq. (3.5):

$$\left| \frac{c^t - c^{t-1}}{c^t} \right| < \varepsilon \tag{3.5}$$

where c^{t-1} is the previous location of the centroid and c^t is the current location of the centroid, t denotes the iteration and ε is a given pre-defined threshold.

3.4 Initial Centroids Selection Principle

In this section, we introduce the mechanism used in k-Means clustering that improves initial centroid selection as well as illustrates how to choose the initial centroids and the reason for this selection. Since we will be using the two new attributes in the CRL file, there is a need to tailor our algorithm to incorporate these changes and optimize its performance on the two-dimensional vector space. In addition, a complex analysis of this algorithm will be provided at the end of this section. As one important prerequisite to carrying through a successful k-Means clustering, the initial centroids selection is critical to the clustering results. This selection needs to be improved upon, which is one of the main goals of our work.

As a widely used technique in many areas such as medicine and computer vision, the original version of k-Means does not require complex computations, however, some limitations still exist. According to [64], those limitations are:

1. User specified the number of cluster k and selected the initial centroid.

2. Clustering results are strongly dependent upon on the initial centroid selection.

3. It may contain the dead unit problem. This occurs if, some units, i.e., centroids, are inappropriately chosen, such as the ones that are far away from the data set compared to the other units; these then become dead units since they may never be updated and cannot properly represent a cluster.

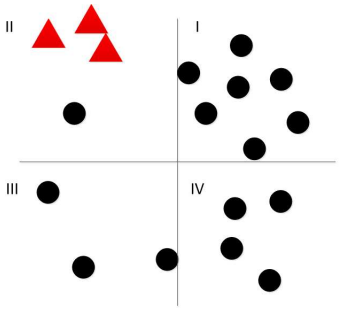

Figure 3.1: An inferior selection of three initial centroids (triangles).

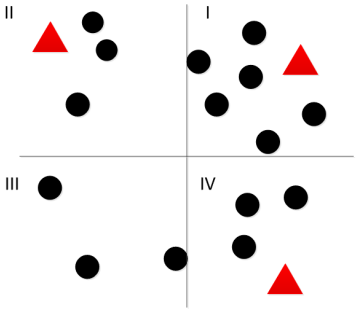

Figure 3.2: A superior initial centroids (triangles) selection.

For the first limitation, we could run the algorithm using different values of k. However, our work mainly deals with the second and third limitation. The improvement is based on two aspects: the distance between newly discovered initial centroids and previous ones, as well as the distribution density of data points in certain zones.

For the problem of distance between current and previous centroids, the original k-Means, as discussed in the previous section, selects the initial centroids randomly without considering their dispensed location. This can increase the chance that some initial centroids may be too close to each other, which might jeopardize the clustering results. In the worse case scenario, when users run the algorithm many times, the initial centroids may be trapped in a certain small area, which would introduce an adverse effect in terms of clustering quality. An example illustration of this challenging situation and a more effective solution are demonstrated respectively in Figure 3.1 and 3.2.

The other issue is the density (also denoted as frequency in the following sections) of the data point distribution. In a vector space containing data points, there are plenty of areas with various densities. Typically, the probability that an area will contain an initial centroid increases as the density of that area increases and, therefore, data point density should be taken into account. For instance, as illustrated in Figure 3.1 and 3.2, Quadrant I of both distribution scenarios show the largest density and, therefore, at least

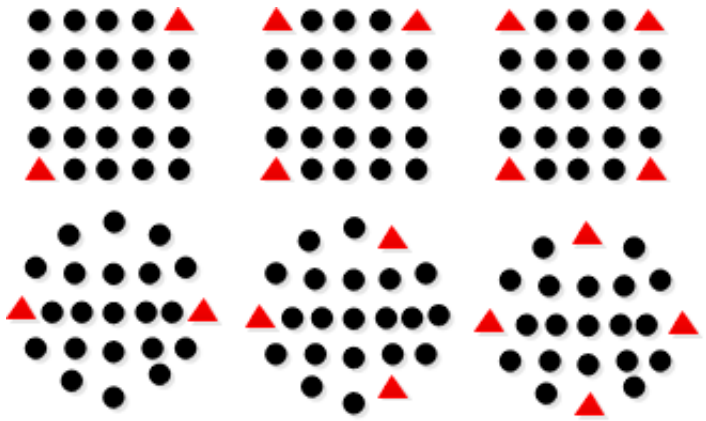

Figure 3.3: An illustration of placing pillars (triangles) that resist the gravity pressure of different shapes of roofs.

one of the initial centroids should be generated within this area.

Our proposed initial centroids selection scheme is inspired by the pillar algorithm for k-Means optimization, proposed in [65], and the modified k-Means algorithm introduced in [66]. This algorithm takes the example of building a structure as an analogy for their work. In this analogy, the roof structure depicts the set of all discrete data points available in all clusters, whereas the pillars represent the initial centroids to be selected from those data points. In order to resist the pressure of a roof structure on a new building, it is wise to place load-bearing pillars as far away as possible from each other, as long as they are within the roof's pressure distribution zone. The placing of two, three, and four pillars with the purpose of withstanding the gravity pressure of two different roof structures in [65] is shown in Figure 3.3. Since the initial centroid placement shares the same concept as the pillar analogy, we propose our algorithm to be based on similar idea. Similar to pillar location selection, initial centroid placement could be done by distributing them evenly and as far away from each other as possible.

Therefore, we propose an enhanced k-Means algorithm with a mechanism where all initial centroids can evenly distributed by taking their distance and density characteristics into account.

3.5 Determining Initial Centroids

In this section, we explain our proposed scheme for initial centroids selection. Firstly, $X = (x_1, x_2, \ldots x_n), x_i \epsilon R^d$ will be used to represent the set of data points. We then partition the two-dimensional vector space into $k * k$ segments, represented by $S = \{S_i | i = 1, 2, \ldots, k^2\}$. The density of each segment is calculated as the quantity of data points in each segment, represented as $F = \{f_i | i = 1, 2, \ldots, k^2\}$.

The segment $S_{max(F)}$, which has the highest frequency, $max\,(F)$, will be selected as the centroids segment, where the first initial centroid μ_1 will be generated in this segment. The $m\left(S_{max(F)}\right)$ is calculated, where $m\,(S) = \{m\,(S_i)\,| i = 1, 2, \ldots, k^2\}$ is used to depict one point in each segment that signifies the geometrical center of all available data points $m\left(S_{max(F)}\right)$ is used as the first initial centroid, denoted as μ_1, and $S_{max(F)}$ is represented as G_1, where $G = \{G_i | i = 1, 2, \ldots, k\}$ is used to denoted some segments that are selected from S and where the future centroid μ_i will be located.

Then, the grand mean, $m(X)$, of all data points in all segments is calculated in the same manner described for finding the geometrical center of one segment. At the same time, the distance between $m\,(X)$ and the μ_1 is measured and this distance is represented as $dis\,(\mu_1, m\,(X))$.

After selecting μ_1, we use a metric called distance metrics $D = \{D_i | i = 1, 2, \ldots, n\}$ in order to spread out the placement of initial centroids. The initial value of D is set as the value of $dis(\mu_1, m(X))$. In order to select the second initial centroid, we use Eq. (3.6) to select next segment $S_j = \{S_j \epsilon S \& S_j \neq S_i\}$, whose frequency is less than the previous centroid segment G_{t-1}.

$$f(S_j) = f(G_{t-1}) - d \tag{3.6}$$

where t is the iteration and d is the difference between the current frequency and previous one, its initial value is set to zero.

If the frequency in the previous iteration $f(G_{t-1})$ is zero, the algorithm is terminated,

otherwise it continues. It should be noted that the quantity q of qualified S_j is unpredictable, e.g., it could be zero, one or more than one. These situations along with their solutions are listed below.

1. $q = 0$, i.e., there is no segment S_j that qualifies, then d is incremented by one until the next segment with the highest frequency is found;

2. $q = 1$, i.e., there is only one segment S_j that qualifies, with S_j assigned to G_t. Then $m(G_t)$ is calculated, which is the initial centroid μ_t for the t-th iteration.

3. $q > 1$, i.e., there are several S_j that qualify equally for selection. Then grand mean $m(S_j)$ is calculated for each of these segments and they are represented as P_i, meaning they are potential initial centroids. $P_i = \{m(S_j) \mid i, j \in [1, n]\}$. Then, we calculate the distance $dis\left(P_i, \mu_{t-1}\right)$ between each potential centroid P_i and the previously assigned centroids μ_{t-1}, denoted as $D_i = \{D_i | i = 1, 2, \ldots, n\}$. Following this, we find the maximum value D_{max} and keep finding the P_i with D_{max}. This P_i is then set as the initial centroid μ_t for the t-th iteration. Therefore this scheme can place the next initial centroid further away from the previously selected ones.

3.6 Formal Description for EKA

In order to better understand the operation of the algorithm proposed, the following description is provided. Let $X = (x_1, x_2, \ldots x_n), x_i \epsilon R^d$ be the set of data points, k be number of clusters, and w be the group width metric. $S = \{S_i | i = 1, 2, \ldots, k^2\}$ is the set of segments that have been partitioned in a two-dimensional vector space, let $F = \{f_i | i = 1, 2, \ldots, k^2\}$ be the frequency in S, $G = \{G_i | i = 1, 2, \ldots, k\}$ be the section where the initial centroids are generated, $\mu = \{\mu_i | i = 1, 2, \ldots, k\}$ be the initial centroids, and $P = \{p_i | i = 1, 2, \ldots, k\}$ be the set of potential initial centroids. $D = \{D_i | i = 1, 2, \ldots, n\}$ denotes the distance metric for each iteration, and $dis(x, y)$ is the function that calculates the distance between data point x and data point y. t represents the current iteration

step, d is the index number used to find the segment in the next iteration that requires an initial centroid to be chosen, and $f(G_i)$ denotes the function that computes the frequency of G_i. Finally, $m(X)$ and $m(G)$ are the grand mean of X and G respectively.

The execution steps are described in Algorithm 3.2.

3.7 Complete Procedure for EKA

Before vehicles and RSUs initialize a conversation with each other, four phases need to be performed during the revocation validation.

1) *Clustering.* In this phase, vehicles and RSUs process the latest CRL file using the two newly added attributes, issued date and credibility, combined with both the k-Means clustering algorithm and the enhanced initial centroids selection scheme in order to efficiently cluster the revocation certificates entries. A sample illustration of the clustering results is shown in Figure 3.4.

2) *Retrieving.* Upon receiving a connection set up request message from other vehicles, receivers will check the certificates contained in the messages and extract all relevant information included in that certificate, i.e., serial number, issue time, and credibility.

3) *Localizing.* Using the credibility and issued date, we can calculate the Euclidean Distance between the data point (i.e., new certificate) and all centroids in order to locate the closest cluster to join.

4) *Verifying.* In this phase, the new data points that join will check all neighboring data points in the recently joined cluster for a match with credibility and issue date. If a match is found, this means that its certificate has been revoked. Otherwise, this data point is not in the CRL and can therefore be trusted.

Algorithm 3.2 Efficient Initial Centroid Selection in K-Means

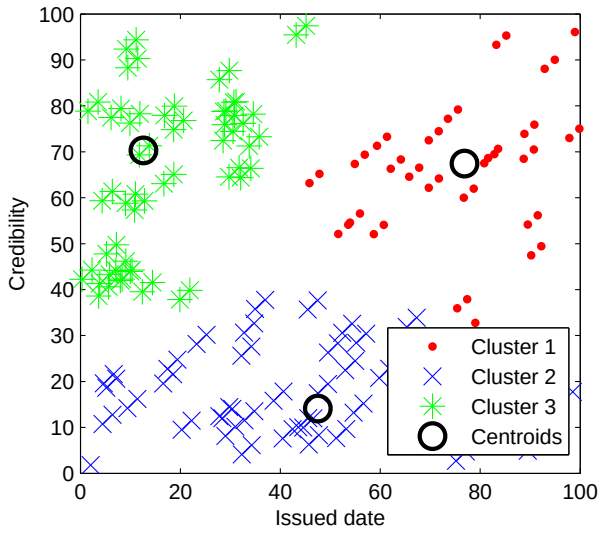

Figure 3.4: An example of clustering results for a CRL, where $n = 100$ and $k = 3$.

Require: Input the number k of cluster centroids.

Ensure: Output k cluster centroids' locations.

1: $S = \emptyset, F = \emptyset, G = \emptyset, \mu = \emptyset, P = \emptyset, D = \emptyset$

2: Calculate $w \leftarrow \{(max(X) - min(X))/k\}$

3: Divide the Vector space into k^2 group with w

4: Assign $S \leftarrow$ segments of vector space

5: $F \leftarrow f(S)$

6: Find $S_i = \{S_i \in S \& F_i = max(F)\}$

7: $G_1 \leftarrow S_i$

8: $\mu_1 \leftarrow m(S_i)$

9: $\mu = \mu \cup \mu_1$

10: Calculate $D \leftarrow dis(\mu_1, m(X))$

11: Set $t = 2$

12: **while** $t \leq k$ **do**

13: **if** $f(G_{t-1} = 0)$ **then**

14: Exit

15: **else**
16: Set $d = 0$
17: { comment: if $q = 0$ }
18: Select $S_j = \{S_j \in S \& f(S_j) = f(G_{t-1}) - d\}$
19: **if** $\neg \exists S_j \in S, f(S_j) = f(G_{t-1}) - d$ **then**
20: $d = d + 1$
21: Go to step 18
22: **else**
23: { comment: if $q = 1$ }
24: **if** $\neg S_j \cdot (S_j \in S \cup (\forall S_i (S_i \in S \rightarrow S_i \neq S_j)))$ **then**
25: $G_t = S_j$
26: Assign $\mu_t \leftarrow m(G_t)$
27: $\mu = \mu \cup \mu_t$
28: Calculate $D \leftarrow dis(\mu_t, \mu_{t-1})$
29: **else**
30: { comment: if $q > 1$ }
31: $\forall S_j = \{S_j \in S \& f(S_j) = f(G_t) - d\}$
32: Calculate $m(S_j)$
33: $P = P \cup m(S_j)$
34: Assign $D \leftarrow max(dis(P, \mu_{t-1}))$
35: Select $p_i = \{p_i \in P \& dis(p_i, \mu_{t-1}) = D\}$
36: Set $\mu_t = p_i$
37: $\mu = \mu \cup \mu_t$
38: **end if**
39: **end if**
40: **end if**
41: $t = t + 1$
42: **end while**

43: Exit

3.8 Summary

This chapter discussed the Efficient K-Means Authentication (EKA) and illustrated its design. We began by introducing the foundation of credibility, i.e., the reputation system. Then the original k-Means clustering, one of the bases of the EKA is presented. Following this, the overall process of initial centroids selection principle is discussed, followed by a formal description of the proposed algorithm used for EKA. Finally, the clustering techniques from data mining are integrated in order to verify whether certificates have been revoked. In this process, two new attributes, credibility and issued date, are used to make certificates cluster-able. In the following chapter, the security analysis of the EKA is shown theoretically and mathematically it can be actualized.

Chapter 4

Security Analysis

Security issues are extremely important in VANETs. As our work focuses on the attacks existing in vehicular communication systems, our concern is the perpetration against messages during communication rather than the physical tampering of vehicles electronics (e.g., against hardware tampering), which will not be covered in this thesis. There are several types of typical attacks in VANETs, which are described briefly in the following sections.

4.1 Preliminaries

In this section, the correctness of the certificate revocation validation scheme is presented. The correctness is based on the preliminaries used in [67]. We employ a similar methodology of analysis as described in [1]. Thus, we improved and simplified the procedure of security analysis in this way.

4.1.1 *Notations*

The symbols in Table 4.1 depict the notations used in this section.

Table 4.1: Notations

Symbol	Notation
$Cert_{A,B,*}$	Certificate issued to B by A.
$CRL_{A,B,*}$	CRL issued to B by A.
$*$	Extra information statement if necessary (optional).
R_i	The i-th RSU.
$PK_{A,*}$, $SK_{A,*}$	The public key and secret key of entity A.
$\partial_{B,*}$	A digital signature signed by entity B.
$Sign(SK_A, M, \sigma_A)$	Signing of message M by *Digital Signature Algorithm* (DSA) with secret key SK_A and attribute set $\underline{\sigma_i}$.
$Verify(PK_A, M, \partial_{A,M})$	Verifying of message M signed by signature algorithms with public key PK_A.
V_j	The j-th vehicle.
$RK_{A,B}$	The returned re-signed key that issued to$B$ by A.
$A_u = \{a_1, a_2, \cdots, a_l\}$	The universal attribute set.
A_i	The attribute set of node A_i.
$enc_h(SK_{CA}, c)$	Encyphering new plaintext c with secret key SK_{CA} and a hash factor$h$.
$dec_h(PK_{CA}, C)$	Decyphering new ciphertext C with public key PK_{CA} and a hash factor$h$.

4.1.2 *Billinear Pairing*

Let there be two finite cyclic groups, $\mathbb{G}$ and $\mathbb{G}_{\mathrm{T}}$, both of them have the same composite order $n = pq$ where p and q are two distinct large primes. Additionally, let $\mathbb{G}_{\mathrm{p}}$ and $\mathbb{G}_{\mathrm{q}}$ be two subgroups of $\mathbb{G}$, their orders are p and q, respectively. A bilinear map e is quipped at both $\mathbb{G}$ and $\mathbb{G}_{\mathrm{T}}$: $\mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_{\mathrm{T}}$. This bilinear mapping has the following properties:

1. Bilinearity: $\forall a, b \in \mathbb{G}, \forall h, k \in \mathbb{Z}_{\mathrm{n}}$, $e\left(a^h, b^k\right) = e\left(ab\right)^{hk}$, the product in the exponent is defined modulo n.

2. Nondegeneracy: $\exists a \in \mathbb{G}$, such that in $\mathbb{G}_{\mathrm{T}}$,$e\left(a, a\right)$ has order n. Let a generate $\mathbb{G}$, then $e\left(a, a\right)$ also be a generator of $\mathbb{G}_{\mathrm{T}}$.

3. Untraceability (ECDLP [1]): Given elements P and Q on an elliptic curve, and a prime q find a number l such that $Q = lP \bmod p$.

A probabilistic algorithm *Gen* takes a secret key $SK_{A,*}$ from an entity A as the input parameter and generates a 9-tuple $(n, p, q, g, u, h, \mathbb{G}, \mathbb{G}_{\mathrm{T}}, e)$, where (g, u) and h are generators of $\mathbb{G}$ and $\mathbb{G}_{\mathrm{q}}$ respectively.

4.1.3 *Attribute Set*

Any node n_j in VANETs can generate a tree structure τ with attributes from A_u. This tree structure can be formally specified as such: Let τ denotes a tree structure rooted at a node r, and each non leaf node x stands for a threshold κ with its value κ_x. The main purpose of the leaf nodes y is representing the attributes $att\left(y\right)$ associated to them. For the non leaf nodes, if one-non leaf node x has s_x sub nodes, the condition of $0 \leq \kappa_x \leq s_x$ must be satisfied.

[1] Elliptic Curve Discrete Logarithm Problem (ECDLP), which is a hard mathematical problem that is believed to be computationally infeasible to solve. Its challenge makes it the basis for the security in cryptography, including the well-known *Elliptic Curve Cryptography* (*ECC*).

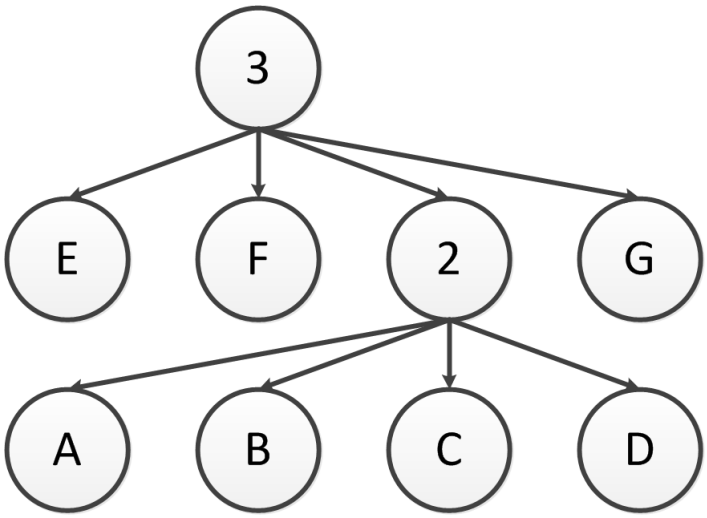

Figure 4.1: The proposed tree structure in [1].

When an attribute set γ has satisfied the following conditions, we define that attribute set γ as a satisfied tree structure τ and use $\tau(\gamma) = 1$ to represent this satisfaction.

Let τ_s be a subtree with its root s be one of the nodes in τ. If this attribute set γ satisfies the tree structure τ_s , $\tau_s(\gamma) = 1$. More specifically,

1. If s is a non leaf node and s has many subnodes z, the quantity of z is at least κ_s, and recursively sub nodes z are satisfied $\tau_z(\gamma) = 1$, then we have $\tau_s(\gamma) = 1$.

2. If x is a leaf node, then $\tau_x(\gamma) = 1$ but only if $att(x) \in \gamma$.

An example is illustrated in Figure 4.1. One eligible attribute set, e.g., $\{B, D, F\}$, is satisfactory to this tree structure.

4.2 System Initialization

CA runs $gen(SK_{CA,*})$ to initialize the system as follows:

1. A bilinear parameter $(n, p, q, g, u, h, \mathbb{G}, \mathbb{G}_{\mathrm{T}}, e)$ is generated, as well as the universal attribute set A_u, where size is l.

2. Suppose that the tree structures that are supported by the attribute-based authentication scheme have the maximum threshold d.

3. CA selects a symmetric encryption algorithm $Sym = \langle Sym.enc, Sym.dec \rangle$, then a secure cryptographic hash function $H : \{0,1\}^* \rightarrow \mathbb{Z}_n^*$, random numbers $\alpha, \delta, a, b, \tilde{t} \in \mathbb{Z}_n^*$ and different number $t_y \in \mathbb{Z}_n^*$ for all the $a_y \in A_u$.

4. CA further selects a redundant attribute set A_r that is indexed from $l+1$ to $l+d-1$.

5. CA additionally computes $\Lambda = e(g,g)^a, \Delta = e(g,u)^\delta, A = g^a, B = g^b, \tilde{T} = g^{\tilde{t}}$, and $T_y = g^{t_y}$ $(1 \leq y \leq l+d-1)$.

6. CA keeps the random number α, a, t_y $(1 \leq y \leq l+d-1)$ and its secrect key $SK_{CA,*}$, and publishes the two system public parameter sets:

$$pubs = (n, g, u, h, \mathbb{G}, \mathbb{G}_{\mathrm{T}}, e, H, Sym,), \tag{4.1}$$

$$pubs' = \left(\Lambda, \Delta, A, B, \tilde{T}, T_y \, (1 \leq y \leq l+d-1)\right). \tag{4.2}$$

If user V_i has an attribute set A_i and registers it to CA, CA will generate a secret key $SK_{V_i,*}$ corresponding to A_i. Specifically, CA chooses random numbers $t, t' \in Z_n$ and a random polynomial $q(x) = \kappa_{d-1}x^{d-1} + \kappa_{d-2}x^{d-2} + \cdots + \kappa_1 x + \delta$ with, degree of $d-1$. CA then calculates $SK_{V_i,*}$ as follows.

$$SK_{V_i,*} = \left\langle K_e, K_d, L, (e_y)_{a_y \in A_i}, (d_y)_{a_y \in A_i \cup A_r} \right\rangle, \tag{4.3}$$

where $K_e = g^\alpha g^{at}$, $K_d = t'$, $L = g^t$, $e_y = T_y^t$ and $d_y = u^{\frac{q(y)}{t'+t_y}}$. CA secretly delivers the secret key $SK_{V_i,*}$ to V_i. Further, pseudonyms and corresponding pseudonym keys will be assigned to nodes by the CA so that nodes are able to periodically change the pseudonyms in the communication in order to preserve their privacy.

4.3 Message Signature and Verification

In message signatures, the vehicle signs the message with the signature algorithm *DSA*.

1. Let a message signer and a receiver be represented by V_i and V_j respectively. $A_i \cup P_i$ indicates V_i's authentication policy τ_i. The threshold value of $A_i \cup P_i$ is κ and Θ_i is an attribute set of τ_i, where τ_i is a single threshold tree structure with maximum supported threshold d. Since $A_i \in A_u$, therefore A_i satisfies $A_u \cup P_i$, which means it is able to find a κ-size attribute set $\Phi_i \subseteq A_i \cap \Theta_i$ and use it in the attribute proof σ_i, which will be generated in next phase.

2. Before transmitting every message M, V_i generates an attribute proof σ_i, then V_i signs M with the signature algorithm $\partial_{V_i,M} = Sign(SK_{v_i}, M, \sigma_i)$. Following this V_i transmits it to the destination. More details is given in the following:

   The generation of the attribute proof σ_i *that performed by node* V_i*:*

   Let the universal attribute set A_u whose size is l. A_r is a redundant attribute set of universal attribute set A_u, A_r indexed from $l+1$ to $l+d-1$. User V_i first chooses a subset $A_{r'} \subseteq A_r$ ($|A_{r'}| = d - \kappa$). Let $A_{r'}$ be $\{a_{l+1}, \ldots, a_{l+d-\kappa}\}$. Then the Lagrangian coefficient $\omega_x = \sum_{w|a_w \in \Psi, w \neq x} \frac{0-w}{x-w}$ for each attribute $a_x \in \Psi = \Phi_i \cup A_{r'}$ is computed by V_i. After that V_i chooses random integers $g, h, K, r_m, r_n, r_x, \in \mathbb{Z}_n^*$ for $a_x \in \Theta_{i,k} \cup A_{r'}$, and $t_x \in \mathbb{Z}_n^*$ for all the $a_x \in A_u$. Additionally V_i computes $T_x = g^{t_x}$ and calculates the generation parameter S_x for $a_x \in \Theta_i \cup A_{r'}$ as follows:

$$S_x = \begin{cases} d^{\omega_x} \cdot h^{r_x} & if\ a_x \in \Psi, \\ h^{r_x} & if\ a_x \in \Theta_i \setminus \Phi_i. \end{cases} \tag{4.4}$$

   The attribute proof is then generated as follows

$$\sigma_i = \left\langle \tau_i, S_m, S_n, (S_x)_{a_x \in \Theta_i \cup A_{r'}}, \varphi_1, \varphi_2 \right\rangle, \tag{4.5}$$

where $S_m = g^{K_d} \cdot h^{r_m}$, $S_n = g^{\frac{1}{K_d+PK_i}} \cdot h^{r_n}$ and $\varphi_1 = S_n^{r_m}\left(g^{PK_i}g^{K_d}\right)^{r_n}$, $\varphi_2 = \prod_{a_x \in \Psi}\left(d_y^{\omega_x}\right)^{r_m} \cdot \prod_{a_x \in \Theta_i \cup A_{r'}}(S_t T_x)^{r_x}$.

3. Upon receiving the message $(M, \partial_{V_i,M,\sigma_i}, Cert_{CA,V_i,*})$ from V_i , the receiver V_j verifies the message by first checking the revocation status of the $Cert_{CA,V_i,*}$ in the latest $CRL_{CA,V_i,*}$. Then it verifies the attribute proof σ_i that is included in $\partial_{V_i,M,\sigma_i}$. More details of the attribute proof σ_i verification will be given in the following:

   Verified the attribute proof σ_i performed by V_j:

   Note that in asymmetric cryptography, the secret key of an entity is the top priority for privacy and maintains a classified status, whereas public keys of entities are accessible. In our case, the public key PK_i of V_i is accessible to V_j. After V_j receives σ_i, V_j will check the following equations to make sure they are both hold.

$$\begin{cases} e\left(S_m g^{PK_i}, S_n\right) \stackrel{?}{=} e\left(g,g\right) e\left(h,\varphi_1\right), \\ \prod_{a_x \in \Theta_i \cup A_{r'}} e\left(S_x, S_m T_y\right) \stackrel{?}{=} \Delta e\left(h,\varphi_2\right). \end{cases} \tag{4.6}$$

   If the equations (4.7) and (4.8) hold and V_i further proves that the A_i' it has matches the private keys PK_i, V_j then confirms that $A_i' = A_i$ andV_i has attributes to satisfy $A_i \cup P_i$. The correctness is proven as follows:

$$\begin{aligned} e\left(S_m g^{PK_i}, S_n\right) &= e\left(g^t \cdot h^{r_m} \cdot g^{PK_i}, g^{\frac{1}{m+PK_i}} \cdot h^{r_n}\right) \\ &= e\left(g,g\right) \cdot e\left(h, \left(g^{\frac{1}{m+PK_i}} \cdot h^{r_n}\right) \cdot \left(g^m \cdot g^{PK_i}\right)^{r_p}\right) \\ &= e\left(g,g\right) \cdot e\left(h, S_p^{r_m}\left(g^{PK_i}g^t\right)^{r_n}\right) \\ &= e\left(g,g\right) \cdot e\left(h,\varphi_1\right), \end{aligned} \tag{4.7}$$

$$\begin{aligned}
\prod_{a_x \in \Theta_i \cup A_{r'}} e\left(S_x, S_m T_x\right) &= \prod_{a_x \in \Psi} e\left(d_x^{\omega_x}, S_m T_x\right) \cdot \prod_{a_x \in \Theta_i \cup A_{r'}} e\left(h^{r_x}, S_m T_x\right) \\
&= \prod_{a_x \in \Psi} e\left(u^{\frac{\omega_x q(x)}{K_d + t_x}}, g^{K_d} \cdot h^{r_m} g^{t_x}\right) \cdot \prod_{a_x \in \Theta_i \cup A_{r'}} e\left(h^{r_x}, S_m T_x\right) \\
&= e\left(g, u\right)^{\delta} \prod_{a_x \in \Psi} e\left(u^{\frac{r_m \omega_x q(x)}{K_d + t_x}}, h\right) \cdot \prod_{a_x \in \Theta_i \cup A_{r'}} e\left(h^{r_x}, S_m T_x\right) \\
&= \Delta \cdot e\left(h, \prod_{a_x \in \Psi} \left(d_x^{\omega_x}\right)^{r_m} \prod_{a_x \in \Theta_i \cup A_{r'}} \left(S_m T_x\right)^{r_x}\right) \\
&= \Delta \cdot e\left(h, \pi_2\right) \qquad (4.8)
\end{aligned}$$

4. If $Verify(PK_{V_i}, M, \partial_{V_i,M,\sigma_i})$ is true, M is accepted; otherwise M is rejected.

Only after the valid ownership of a certificate is shown by the sender and the verification of the revocation status is completed can this message be accepted and the authenticity of the message be guaranteed by using a signature algorithm. Thus, common attacks can be prevented and non-repudiation can be achieved as well.

4.4 Certificate Revocation

When a vehicle V_i is compromised, its certificate is added as an entry in the CRL. In order to do so, the following steps are followed:

1. The CA sends the pseudo identity information of the revoked vehicle certificate $Cert_{CA,V_i,T}$ to all RSUs. In addition to revoking V_i thoroughly, the CA also prevents it from accessing vehicular communications for a certain revocation period, such as between the current time window TW_T and a future one $TW_{T'}$ $(T' \in (T, C])$, where C is the time when the next CRL released. This prevention is carried out beginning with CA calculating $S_{1,T} = h^T\left(SD_1\right)$, $S_{2,C-T+1} = h^{C-T+1}\left(SD_2\right)$, where SD_i is the seed of the hash function. It then sends the pseudo identity information of the revoked certificates $\langle T, T', S_1, S_{2,C-T+1}\rangle$ to all RSUs.

2. After receiving $Cert_{CA,V_i,T}$ and $\langle T, T', S_1, S_{2,C-T+1}\rangle$, an R_x calculates the pseudo

identities $PID_k(k \in [T, T'])$ of revoked certificates, where

$$\begin{cases} S_{1,k} = h^{k-T}\left(S_1, T\right), \\ S_{2,C-k+1} = h^{T'-k}\left(S_{2,C-T'+1}\right) \\ PID_k = h\left(S_{1,k} \parallel S_{2,C-k+1}\right) \end{cases} \tag{4.9}$$

each RSU R_x adds the related information PID_k to its local CRL. Therefore, the revoked certificate $Cert_{CA,V_i,T}$ would no longer be able to request a re-signing of the certificate from RSUs. All R_xs will then returning a confirmation message $(M, SK_{R_x}, \partial_{CA,V_i})$ to the CA. The R_x will then determines which clusters of CRL the new revoked certificate $Cert_{CA,V_i,T}$ will add based on the revoked certificate's credibility and issued date.

3. Upon receiving the confirmation from all RSUs, the CA adds the information of the revoked certificate to $CRL_{CA,V}$, which is shared among all vehicles at a later point. The CA then calculates $S_{3,(T-1)*L_w+1} = h^{(T-1)*L_w+1}\left(SD_3\right)$, and $S_{4,(C-k)*L_w+1} = h^{4,(C-k)*L_w+1}\left(SD_4\right)$, where L_w is the length of the time window. It then adds the pseudo identity information of the revoked pseudonymous certificate

$$\left\langle (T-1)*L_w, k*L_w, S_{3,(T-1)*L_w+1}, S_{4,(C-k)*L_w+1} \right\rangle \tag{4.10}$$

to $CRL_{CA,V}$. At the same time, each RSU R_x will broadcast

$$\left\langle M, \partial_{R_x,V_j}, (T-1)*L_w, k*L_w, S_{3,(T-1)*L_w+1}, S_{4,(C-k)*L_w+1} \right\rangle \tag{4.11}$$

to all vehicles $V_j, V_j \neq V_i$ that are within its covered area.

4. When vehicle V_j receives $Cert_{CA,V_i,T}$ and a vehicle calculates these pseudo identities $PID_j\,(j \in (n-1]*L_w, k*L_w)$of revoked pseudonymous certificates, where

$$\begin{cases} S_{3,j} = h^{j-(n-1)\times L_w-1}\left(S_{3,(n-1)\times L_w+1}\right) \\ S_{4,C\times L_w-j+1} = h^{k\times L_w-j}\left(S_{4,(C-k)\times L_w+1}\right) \\ PID_j = h\left(S_{3,j} \parallel S_{4,C\times L_w-j+1}\right) \end{cases} \tag{4.12}$$

PID_j will be add the revoked vehicle certificate to the local current CRL_{CA,V_j} and determine which clusters the new revoked certificate will add.

5. When it is time for the next CRL update, each vehicle V_j will receive a complete version of $CRL_{A,V}$, which includes all the recently revoked certificates. After receiving the $CRL_{A,V}$, a vehicle updates it local CRL in its entirety.

The revoked certificate privacy could be preserved as anonymous channel that is secure from other vehicles and can be used to communicate private information between the RSU and the CA. Each vehicle can have the latest certificate revocation update from the RSU as long as they are within the RSU coverage area.

The R_x can distribute the revoked certificate message $Cert_{CA,V_i,T}$ using moving vehicles that are within its covered area in an epidemic manner. At first, RSUs broadcast revoked certificate messages $(M, Cert_{CA,V_i,T})$ and any V_i receiving $(M, Cert_{CA,V_i,T})$ is considered infected. Each infected vehicle then continuously infects all vehicles it passes by. By using the steps distribution of the revoked certificate message $Cert_{CA,V_i,T}$ can be achieved.

4.5 Certificate Re-signing

Vehicle V_i passes by an R_x and sends their own certificates $Cert_{CA,V_i,*}$ to R_x in order to have their certificates periodically re-signed. This re-signed signature along with timestamps can prove that the certificate is current. Valid certificates can get re-signed simply by sending a request to the RSU when passing by it. If the certificate is not revoked, R_x timestamps the certificate to denote it is valid and returns it to the V_i . Otherwise the RSU will refuse to re-sign it. The current status of the signature denotes the validity of the certificate. Given that T is the current time, the entire process is described in the following steps:

1. Initially, CA generates a secret key SK_i to user V_i who has an attribute set A_i. Then

a tree structure is generated and, a leaf node is chosen by the CA and associated with SK_i. If this leaf node has previously been associated with SK_j for another node V_j, this assignment cannot be completed. Next, the CA generates a node set $P_u(SK_i)$ and chooses several integers $\hat{t}, \tilde{t}, r_d, y \in \mathbb{Z}_{\mathrm{q}}^*$ for $y \in P_u(SK_i)$. By using its own secret key SK_{CA}, the CA generates the secret key SK_i for V_i,

$$SK_i = \left\langle K, L, \{K_x\}_{x \in A_i}, \{D_y, d_y\}_{y \in P_u(SK_{CA})} \right\rangle, \tag{4.13}$$

where $K = g^a g^{a\hat{t}} g^{b\bar{t}}$, $L = g^{\hat{t}}$, $K_x = h_x^{\hat{t}}$, $D_y = B^{a_y d + \tilde{t}} H(d)^{r_{d,y}}$, $d_y = g^{r_{d,y}}$.

2. At the end of each time slot t, V_i send its own certificate with secret key SK_i to R_x, then R_x generates a re-signature key corresponding to the signing certificate $Cert_{CA,V_i,T'}$, where $T' > T$ as an exponent in $\mathbb{Z}_q$.

3. R_x periodically broadcasts $(M, Cert_{CA,R_x})$ to remind every incoming vehicle that they are entering the area covered by R_x. Also at this time, the CA prepares a revocation list $CRL_{T'}$ including all revoked certificates.

4. When V_i receives $Cert_{CA,R_x}$, it sends the request message $(t_{stamp} Cert_{CA,V_i,T})$ to R_x, where t_{stamp} is the issued time timestamp.

5. When R_x receive this message, If t_{stamp} is current, i.e., still within the valid period, and $Cert_{CA,V_i}$ is not revoked, R_x sends the re-signature key RK_{R_x,V_i} and $t_{stamp} Cert_{CA,V_i,T'}$ back to V_i. Following this, R_x records the current time, T', and a certificate $< T', t_{stamp} Cert_{CA,V_i,T'} >$ is created. The CA also creates a certificate set $K_i(CRL_{T'})$ corresponding to $Cert_{CA,V_i}$ from τ. The CA then chooses random exponents $r_{t,y} \in \mathbb{Z}_q$, and outputs the updated information τ_t,

$$\tau_t = \left\langle \{E_y e_y\}_{y \in K_i(CRL_{T'})} \right\rangle, \tag{4.14}$$

where $E_y = D^{a_x t} H(t)^{r_{t,y}}$ and $e_y = g^{r_{t,y}}$.

6. V_i checks the RK_{R_x,V_i} for the presence of the $t_{stamp} Cert_{CA,V_i,T'}$.

A malicious vehicle may attempt to generate a certificate with an invalid identity to prevent itself from being tracked by the CA. Since the RSU has signed the message via $Sign(SK_{V_i}, M)$, the vehicle cannot forge a certificate due to other vehicles, SK_{V_i} being confidential.

4.6 CRL Issuing

In CRL issuing, the CA issues a CRL_{CA,R_x} to the RSU R_x, which is positioned in one assigned area. This process and the process of CA issuing to vehicle V_i can be described as follows:

1. A hash number h is calculated by SHA-1 cryptographic hash function with a key that is the MAC address of the receiver R_x network interface controller.

2. The CA then sets the secret key $SK_{CA,*} = h$ and generates the public key PK_{R_x}. Let *crl* denote the new CRL, which is ready for distribution, M is a $\mu \times \theta$ matrix, and the i-th row of M is denoted as M_i. Let ρ be the mapping from $\{1, 2, \cdots, \mu\}$ to the attribute index $\{1, 2, \cdots, l\}$. Then the CA will encrypt the *crl* by going through the following procedures:

   Encryption performed by CA:

   The CA chooses a random vector $\vec{v} = (v_1 = s, v_2, \cdots, v_\theta) \in \mathbb{Z}_{\mathrm{n}}^{\theta}$, random numbers $v_1, \cdots, v_\theta \in \mathbb{Z}_n$ and its own secret key $SK_{CA} \in C_T$. For $1 \leq i \leq \mu$, the CA calculates $\lambda_i = \vec{v} \cdot M_i$. The ciphertext is $C = \left\langle C, C', C_s, (C_i, D_i)_{1 \leq i \leq \mu} \right\rangle$, where $C = SK_{CA} \cdot \lambda^s$, $C' = g^s$,$C_i = A^{\lambda_i} T_{\rho(i)}^{-r_x}$, $D_i = g^{r_i}$ and $C_s = enc_h(SK_{CA}, crl)$.

3. The signature ∂_{CA,R_x} is generated by the CA with ciphertext C, where $\partial_{CA,R_X} = Sign(SK_{CA,*}, PK_{R_x})$. Both ∂_{CA,R_x} and ciphertext C are distributed.

4. After going through the following operations, the CA will be defined as successful and will deliver PK_{CA}, PK_{R_X} and $CRL_{CA,R_x,*} = (PK_{R_x}, \partial_{CA,R_x})$ to R_x. The

mapping θ between the R_x and $CRL_{CA,R_x,*}$ is then stored by R_x.

Decryption performed by V_j:

V_j receives the ciphertext C from CA. V_j has $E_j = \left\langle K_e, K_d, L, (e_y, d_y)_{a_i \in A_j} \right\rangle$and let $\tau = \left\{ y |\ \mathrm{a}_{\rho(\mathrm{i})} \in \mathrm{A_j} \right\}$. It is computationally feasible to find $\{\omega_i \in \mathbb{Z}_\mathrm{n}\}_{i\in\tau}$, so that $\sum_{i\in\tau} \omega_i \lambda_i = s$.$V_j$ then calculates $\frac{C \cdot \prod_{i\in\tau} \left(e(C_i, L) \cdot e\left(D_i, e_{\rho(i)}\right)\right)^{\omega_i}}{e(C', K_e)} = PK_{CA}$ and $dec_h\left(PK_{CA}, C_s\right) = crl$.

The correctness of the decryption algorithm is proven below:

$$\begin{aligned}
C \cdot \prod_{i\in I} \left(e\left(C_i, L\right) \cdot e\left(D_i, e_{\rho(i)}\right)\right)^{\omega_i} &= \frac{PK_{CA} \cdot \lambda^s \cdot \prod_{i\in I} \left(e\left(C_i, L\right) \cdot e\left(D_i, e_{\rho(i)}\right)\right)^{\omega_i}}{e\left(g^s, g^\alpha g^{at}\right)} \\
&= \frac{PK_{CA} \cdot \prod_{i\in I} \left(e\left(A^{\lambda_i} T^{-r_i}, g^t\right) \cdot e\left(g^{r_i}, T^t_{\rho(i)}\right)\right)^{\omega_i}}{e\left(g^s, g^{at}\right)} \\
&= \frac{PK_{CA} \cdot \prod_{i\in I} \left(e\left(A^{\lambda_i}, g^t\right)\right)^{\omega_i}}{e\left(g^s, g^{at}\right)} \\
&= \frac{PK_{CA} \cdot e\left(A^{\lambda_i}, g^t\right)^s}{e\left(g^s, g^{at}\right)} \\
&= PK_{CA}
\end{aligned} \tag{4.15}$$

5. R_x can verify its own CRL $CRL_{CA,R_x,*}$ using $Verify(PK_{CA}, PK_{R_x}, \partial_{CA,R_x})$.

Since the signature algorithm is applied, the RSUs CRL_{CA,R_x} are proven to be authentic. During the issuing, the CA associates the MAC address of the receiver R_x or V_i to the digital certificate so as to restrict the user of a specific digital certificate. This prevents masqueraders from pretending to be other legitimate nodes since the MAC address of masqueraders cannot be identical to that of the R_x or V_i. In addition, vehicle receivers can verify the R_x digital certificates if necessary.

By binding the unique MAC address to the digital certificate, we can ensure that other compromised RSU nodes cannot pretend to be the certificate holder using this certificate.

4.7 Summary

In this chapter, further details of security analysis regarding the application of EKA have been presented. In order to broadcast messages, CAs could be used to authenticate nodes and provide some nodes with a privilege of broadcasting based on their authorization level, which is associated with their certificate's credibility rating. Further, for the purpose of secure broadcasting privilege related to safety-critical events, those privileges are only given to nodes with a higher level of authorization due to their high credibility. These safety-critical related events (e.g., huge fallen rock detected on the road ahead) should be broadcaster by the CA or high level nodes only, while nodes with lower levels can share non-critical information. In addition, senders need to sign each message to be broadcaster as the source, and therefore, the CA and other nodes can easily trace the originator . All certificates that are listed in the CRLs can be disqualified from broadcasting privilege because they have been revoked. This scheme assures that broadcasting privileges are not being abused by malicious nodes or nodes whose certificates have been revoked.

Several communication technologies are compatible to work with VANETs (e.g., UMTS' Terrestrial Radio Access Time Division Duplex, Wi-MAX, Wi-Fi, and ZigBee). The Long-Term Evolution (LTE) standard that is commonly considered as a 4G technology in Global System for Mobile communications (GSM) has been considered for its flexibility. In other words, it could be tailored to be used in a wide range of existing frequencies and spectra. Obviously, it will be more secure to have multiple communication systems rather than just one, so that vehicular communication systems could use different communication technologies. For example, when a jamming type of attack is performed, communicating vehicles can switch to another communication technology in order to successfully send their information.

Chapter 5

Performance Evaluations

Our simulations were performed on the network simulator-2 [68] using one mobility model that that is similar to the *Manhattan Mobility Model* [69] in defining the movement of vehicles. The radio-propagation model was based on the two-ray ground reflection model using the IEEE 802.11 standards for MAC address models, along with a one-hop transmissions model, i.e., the DumbAgent model. Moreover, simulations employed the omnidirectional antenna model using a 300 meter node communication range. Table 5.1 shows the detailed environment settings for the implemented simulations.

The two node authentication algorithms are evaluated in this section. These algorithms searches linearly through the CRLs, that is, authentication scheme goes through the CRL file in a linear way and searches for a match; and the other algorithm follows the algorithm proposed in EKA. The elapsed time for each authentication process is recorded for comparison purposes.

5.1 Simulation Scenarios

In this section, we briefly demonstrate the scenarios we used to perform extensive sets of simulation experiments. After describing our scenarios, we will present the metrics and simulation results.

Table 5.1: Simulation environments

Environments	Settings
Simulator	Network simulator-2
Mobility model	Manhattan Mobility Model
Radio-propagation model	Two-ray ground reflection model
MAC address model	IEEE 802.11 standards
Transmissions model	DumbAgent model
Antenna model	Omnidirectional
communication range	300 m

5.1.1 Experimental Environment

The mobility zone for an simulation was a square of $l * l\ m^2$. Within this simulation zone, v vehicle nodes were randomly deployed. The vehicle nodes moved using a mobility model that similar to the famous *Manhattan Mobility Model*, with vehicle nodes being provided with random destinations and a speed that was of uniform distribution from zero to the pre-specified maximum velocity. Once the vehicle node arrived at the intended destination, the nodes randomly chose the next destination and traveled towards it with a different speed. After simulation time T, the simulation was terminated. An illustration of our mobility model is shown in Figure 5.1.

5.1.2 Authentication Process

During the simulation, the vehicle nodes broadcast messages that contain their serial number, credibility and issued date to other peer vehicles. Once other nodes came into communication range of the broadcasting vehicle, they would receive the messages and extract all the information in order to conduct authentication by verifying the CRL file using two searching schemes: linear searching and EKA. The linear search is represented as *l-searching* in the following comparison whereas the EKA search scheme is denoted as

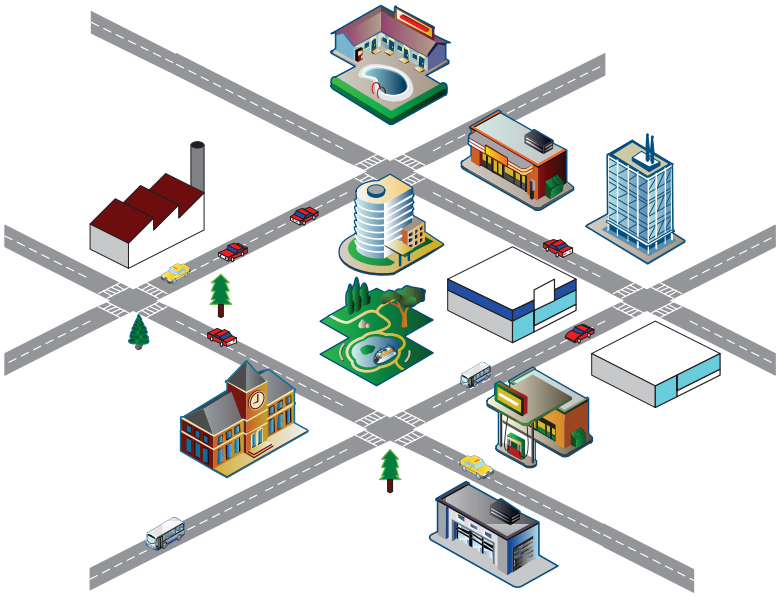

Figure 5.1: An illustration of our mobility model.

k-searching. Furthermore, the time that was needed to conduct clustering in the k-Means clustering algorithm was also recorded, the sum of this and the *k-searching* consisted of the total execution time for the EKA, which is denoted as *k-full*.

The CRL may carry thousands of revoked certificate records and we use s to represent the size of CRL, which also denotes the number of records that are listed in the CRL k denotes the number of clusters. The serial number, credibility and issued date of each entry in the CRL were generated randomly and the attribute values of some vehicles and RSU certificates were assigned based on some of the randomly select records in order to act as revoked certificates.

5.1.3 Parameters and Simulations

Table 5.2 shows in detail the parameter settings for the implemented simulations. Two search schemes (linear searching and EKA) were evaluated for performance comparisons.

Table 5.2: Simulation parameters

Parameter	Value
Running time(t)	$1000, 1500, 2000, 2500, 3000s$
Number of nodes(n)	$100, 200, 300, 400, 500$
Size of CRL(s)	$1000, 3000, 5000, 7000$
Number of clusters(k)	$5, 7, 10, 13, 15$
Density of mobility zone(d)	$25, 50, 75, 100, 125$
Length of simulation zone(l)	$1000m, 2000m$

The EKA was modified to perform authentication and five sets of simulations were implemented for evaluation. All of these evaluations were performed in the previously defined environment. The five sets of experiments performed were done using the following varying values:

1. *Running time of simulation* (t): Two search schemes were measured for the purpose of comparison using a varied simulation time ranging from 1000s to 3000s, with increments of 500s.

2. *Number of vehicle nodes and RSU nodes* (n): The evaluation compared the elapsed time between different numbers of nodes in the simulation. The number of nodes varied from 100 to 500, with increments of 100.

3. *Size of CRL* (s): The evaluation was undertaken to analyze the effect of the size of the CRL on the performance of the two search schemes being evaluated. The size, s, was varied from 1000 to 7000, with increments of 2000.

4. *Number of clusters in k-Means clustering* (k): In this set of experiments, the number of clusters was varied from 5 to 15. This was used to visualize the effect of increasing the number of clusters on the authentication process in order to select a suitable number of clusters.

5. *Density of mobility zone* (d): This set of experiments compared the impact of mobility density on the performance of the two evaluated mechanisms. The density d was calculated using the equation $d = \frac{n}{l*l}$, where l is the length of the simulation zone.

5.2 Performance Metrics

Two metrics were utilized to assess the performance of our authentication mechanism.

1. *CRL Verification Latency* (CVL): The total time needed for all nodes to complete the CRL verification.

2. *Communication Overhead*: The average number of messages broadcast by all RSUs and vehicle nodes during execution time.

5.3 Simulation Results and Analysis

In this section, we will present the results of a set of extensive simulations and a comprehensive analysis of these results will also be provided. In each case, we sample enough experiment of random vehicular movements to put our results in a 95 percent confidence interval.

5.3.1 Impact of Different Simulation Run Times

Table 5.3 compares searching execution time for the linear searching scheme and the EKA with varying simulation run times. The increase in of the simulation run time increases the execution time of both search schemes. However, the execution time of *k-searching* and *k-full* remained under 10s in all five simulations. On the contrary, the execution time increased dramatically with the linear search scheme.

Table 5.3: Impact of various simulation run times to CVL.

Run time(t)	Communication Overhead	Average CVL (s) *(C.I.)*		
		l-searching	*k-searching*	*k-full*
1000	119204	28.67 (1.07)	0.50 (0.13)	2.52 (0.14)
1500	178932	50.80 (2.31)	0.88 (0.18)	3.14 (0.20)
2000	240198	67.85 (5.27)	1.13 (0.22)	3.15 (0.35)
2500	305368	77.12 (4.82)	1.29 (0.27)	3.17 (0.58)
3000	366120	83.44 (3.65)	1.39 (0.31)	3.18 (0.23)
where $n = 100, k = 3, s = 1000, l = 1000m$. C.I.: confidence interval.				

The results demonstrate that the EKA promotes the authentication latency by reducing the number of entries in the CRL that the sender's certificate needs to be verified with.

Once the CRL clustering is complete, the EKA executions overhead in our mechanism only includes *k-searching*. The *k-searching* grew as the simulation run time increased, whereas the *k-full* did not vary by any significant degree since the execution time for the CRL clustering was determined provided the number of revoked certificates was set and the number of clusters was specified. The reasons that the increase of simulation run time increased the needed execution time for *k-searching* is that the increase to the simulation run time produced more broadcast messages; thus so the time for each node to verify the all messages it received was greater than before.

5.3.2 Impact of Different Quantities of Nodes

Five different thresholds for the number of vehicles and RSU nodes were evaluated in the simulation under the same size of simulation zone (the number was varied from 100 to 500, with increments of 100). The authentication latency for both authentication schemes

Table 5.4: Impact of different quantities of nodes to average CVL.

Quantity of nodes(n)	Communication Overhead	Average CVL (s) *(C.I.)*		
		l-searching	*k-searching*	*k-full*
100	41854	7.15 (2.67)	0.09 (0.01)	3.56 (0.63)
200	157422	26.45 (4.70)	0.36 (0.09)	6.70 (0.85)
300	334563	61.94 (6.85)	0.76 (0.07)	10.52 (1.32)
400	561190	95.50 (3.21)	1.27 (0.12)	14.29 (2.42)
500	819168	138.74 (4.21)	1.84 (0.31)	17.31 (1.13)
where $k = 10, s = 1000, l = 2000m, t = 1000s$. C.I.: confidence interval.				

was increased following the growth of the threshold, yet the authentication latency of EKA grew only slightly compared to the linear search scheme. In addition, the ratio of CVL for the linear and EKA searching scheme was shown in Table 5.4, which depicts the ratio rose rapidly. When the number of nodes was over 200, the ratio was almost 4, which means the execution time of linear search scheme is almost four times longer than that of the EKA. The ratio was further improved to about 10 if the threshold was 500.

The time needed for authentication latency was prolonged because both search schemes needed to process more messages for the purpose of authentication. The node overhead was related to the execution time so setting a higher threshold also introduced an increase in the number of messages. Based on the above results, the advantages of EKA are obvious when faced with an increasing number of vehicles and RSUs nodes.

5.3.3 Impact of Various CRL Size

Table 5.5 illustrates the performance of the two mechanisms using three CRLs that have sizes that differ from each other. Both the authentication latency and communication overhead increased with a larger size of CRL. The reason for this is that with a larger

Table 5.5: Impact of various CRL size to average CVL.

Size of CRL(s)	Communication Overhead	Average CVL (s) *(C.I.)*		
		l-searching	*k-searching*	*k-full*
1000	157423	26.45 (5.59)	0.36 (0.07)	6.70 (1.43)
3000	157590	68.96 (4.91)	0.50 (0.05)	26.39 (4.29)
5000	157920	111.15 (2.31)	0.64 (0.10)	50.41 (3.71)
7000	157821	130.96 (13.58)	0.79 (0.21)	68.08 (4.74)
where $k = 10, n = 200, l = 2000m, t = 1000s$. C.I.: confidence interval.				

size of CRL, the linear search scheme needed to verifying more entries in the CRL file, thus the nodes spent more time on the presence of verification, This reason also applied to the EKA. The needed authentication overhead for EKA was also enlarged due to increasing the size of the CRL. As a result of the increased CRL size, the EKA needed to deal with more entries and have them clustered. Therefore the authentication latency was prolonged.

5.3.4 Impact of Various Quantities of Clusters

As shown in Table 5.6, the average CVL for the *l-searching* and *k-searching* were little affected by the different numbers of clusters. However, in the case of the *k-full*, unlike the others, increasing the number of clusters helped to increase the latency. The reason for this is that more clusters cause nodes to take more time to find which centroid cluster is closest to them; this is shown in step 6 to step 10 in Algorithm 3.1 in Chapter 3.

Table 5.6: Impact of different quantity of cluster to average CVL.

Quantity of cluster(k)	Communication Overhead	Average CVL (s) *(C.I.)*		
		l-searching	*k-searching*	*k-full*
5	41854	7.65 (1.59)	0.09 (0.02)	1.40 (0.17)
7	41854	7.61 (1.17)	0.10 (0.01)	2.27 (0.05)
10	41853	7.61 (0.23)	0.10 (0.02)	3.39 (1.32)
13	41869	7.57 (1.37)	0.12 (0.02)	4.25 (0.17)
15	41909	7.61 (0.74)	0.11 (0.01)	4.59 (0.04)
where $n = 200, s = 1000, l = 2000m, t = 1000s$. C.I.: confidence interval.				

5.3.5 Impact of Various Density of Mobility Zone

As a result of increasing the density of the simulation zone, the average CVL of *l-searching* steeply rose from 7.15s to 138.74s, as displayed in Table 5.7. The average CVL for both *k-searching* and *k-full* also increased. However, with EKA, the speed at which the latency increased can be effectively slowed down. The average CVL for *k-full* rose from 3.56s to 17.31s when the number of nodes reached 500. The linear search scheme involved more execution time to tolerate the variation of the mobility zone density. With the density of 125 n/km^2, the average CVL of *l-searching* was increased to a level that was almost 10 times longer than that of *k-full*.

Table 5.7: Impact of various mobility zone density to average CVL.

Density (n/km^2)	Communication Overhead	Average CVL (s) *(C.I.)*		
		l-searching	*k-searching*	*k-full*
25	41854	7.15 (1.32)	0.09 (0.01)	3.56 (0.21)
50	157422	26.45 (3.97)	0.36 (0.01)	6.70 (1.46)
75	334563	61.94 (4.16)	0.76 (0.06)	10.52 (1.85)
100	561190	95.50 (3.24)	1.27 (0.19)	14.29 (1.64)
125	819168	138.74 (8.10)	1.84 (0.05)	17.31 (2.65)
where $l = 2000m, s = 1000, k = 10, t = 1000s$. C.I.: confidence interval.				

Chapter 6

Conclusions and Future Work

Due to the unique characteristics of VANETs, such as the velocity of vehicles, unpredictably large quantities of either vehicles or digital certificates that they possess, and the high number of certificates issued and revoked, traditional certificate revocation status validation schemes are not suitable for authentication for VANETs.

In this thesis, an efficient certificate revocation status validation scheme, i.e., EKA, has been presented to provide reliable, secure and rapid certificate-based authentication for VANETs. In this chapter, we will provide a summary of our thesis work as well as pointing to possible directions for future research.

6.1 Conclusions

For many years, authentication has been a field that academia has concerned itself with. Many researchers have looked at different techniques that may facilitate and increase the efficiency of verifying certificates. Typically, these techniques have tried to verify certificates using CAs and have aimed to reduce the usage of bandwidth during the process of verification. Those schemes work well in wired environments, however, are less useful in ad hoc networks due to the characteristics of the dynamic topology. This is especially evident in VANETs, where the speed of nodes' movement is often very high.

This thesis has presented some important new additions to certificates in the context of VANETs and has proposed a novel certification verification scheme that is suitable for use in VANETs.

At the outset of this thesis, we explained why vehicular communication requires the assurance of a peer's credibility to be delivered in a quick manner and why this problem needs a specific approach. To provide more details, the available studies and research in the field's literature were reviewed with regards to five topics: wireless networks, wireless ad hoc networks, VANETs, authentication techniques in PKI and data mining. Some developing aspects of each topic was discussed, such as wireless mesh network in wireless ad hoc networks, and so on. Furthermore, with regard to authentication, we explained fully the detailed structure of a X.509 digital certificate, the main responsibility of CAs, the format in which data is represented in a CRL file. With these overviews, a broader view of the work completed in this field as it relates to this thesis was accomplished.

Next, a broad introduction to reputation management was provided. Serving as the bases of credibility, the definition of reputation and most common types of reputation models was presented, followed by reputation models used in MANETs and VANETs, respectively. The challenges of reputation models in ad hoc networks were also discussed in this section. Following this, different approaches to by which to calculate the value of credibility were provided. These different approaches were tailored to fit the requirements of VANETs, and were also represented in a form that can be utilized to calculate different credibility status of certificates.

The thesis then discussed two novel attributes that are suggested as an addition to certificate standards, these being credibility and issued date, and explained the reasons and benefits for this addition. The thesis then went on to describe the original k-Means clustering algorithm as well as an algorithm that depicts the operation of k-means clustering procedures. Further to this, the limitations of the k-means clustering algorithm and the importance of initial centroid selection were discussed. We then proposed an enhanced initial centroid selection algorithm that optimizes the selection process. We

followed this by discussing what modifications need to be introduced into the original k-Means algorithm and showed how to adopt the clustering concept into the certificate revocation status.

Subsequently, we followed the previous chapter with a number of mathematical security analyses that demonstrated proposed schemes to increase with the security of VANETs. The security analysis was mainly focused on four aspect of certificate revocation validation: message signing and verifying, CRLs issuing, certificates re-signing and certification revocation.

Finally, we evaluated the performance of our proposed approach in terms of some important network performance metrics; namely, verification latency and communication overhead. First, we briefly introduced the simulation scenario, then the process of authentication and the various parameters for five sets of experiments were given. These experiments were performed using the following varying values: run time of simulation, number of vehicle nodes and RSU nodes, size of CRL, number of clusters in k-Means clustering, density of mobility zone. We then proceeded to define the two network performance metrics that were measured in order to make a comparison. Eventually, with the help of the simulation results, we successfully proved that adopting the clustering concept in revocation status validation is suitable for the considered problems and ultimately performs better than the traditional linear CRL search.

6.2 Future Work

Direction for future research is planned as follows:

- *Support for distribution of clustering information*

We plan to develop this scheme further in future work. More specifically, we intend on exploring and verifying the feasibility of different ways in which the CA can distribute clustering information along with the CRLs. Eventually, these can be more economic in terms of time, since each vehicle will have the clustering information beforehand.

- *Development of an offline cryptographic scheme*

Another interesting challeng, which we currently have not resolved, is to develop an offline cryptographic scheme that will be able to exclude specific nodes from participation when transmitting and receiving messages. For instance, if any node could not prove its knowledge of a group's confidential information, which could be renewable or revocable in an efficient fashion, or if the credibility rating values are lower than a specific threshold, the ideal scheme would be able to prevent them from participating in the communication. This ideal scheme can be designed as an "elimination" scheme, in order to only allow for legitimate nodes to form a secure and trusted group.

Index

Bibliography

[1] X. Liang, M. Barua, R. Lu, X. Lin, and X. S. Shen, "Healthshare: Achieving secure and privacy-preserving health information sharing through health social networks," *Computer Communications*, no. 0, pp. –, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366412000102

[2] A. Boukerche, Ed., *Algorithms and Protocols for Wireless, Mobile Ad Hoc Networks*, 1st ed. Wiley-IEEE Press, Nov. 2008.

[3] A. Boukerche, *Algorithms and protocols for wireless sensor networks*. Wiley, 2009.

[4] T. Mak, K. Laberteaux, R. Sengupta, and M. Ergen, "Multichannel medium access control for dedicated short-range communications," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 1, pp. 349 –366, jan. 2009.

[5] VeriSign, "VeriSign CRL," http://crl.verisign.com/. [Online]. Available: http://crl.verisign.com/

[6] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894 –914, may 2009.

[7] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *Communications Surveys Tutorials, IEEE*, vol. 11, no. 1, pp. 116 –130, quarter 2009.

[8] C. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. Shellhammer, and W. Caldwell, "Ieee 802.22: The first cognitive radio wireless regional area network standard," *Communications Magazine, IEEE*, vol. 47, no. 1, pp. 130 –138, january 2009.

[9] P. Li, C. Zhang, and Y. Fang, "Capacity and delay of hybrid wireless broadband access networks," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 2, pp. 117 –125, february 2009.

[10] J. Yao, "Microwave photonics," *Lightwave Technology, Journal of*, vol. 27, no. 3, pp. 314 –335, feb.1, 2009.

[11] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, ser. WSNA '02. New York, NY, USA: ACM, 2002, pp. 88–97. [Online]. Available: http://doi.acm.org/10.1145/570738.570751

[12] M. Chitnis, Y. Liang, J. Y. Zheng, P. Pagano, and G. Lipari, "Wireless line sensor network for distributed visual surveillance," in *Proceedings of the 6th ACM symposium on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, ser. PE-WASUN '09. New York, NY, USA: ACM, 2009, pp. 71–78. [Online]. Available: http://doi.acm.org/10.1145/1641876.1641890

[13] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient routing in intermittently connected mobile networks: The Multiple-Copy case," *Networking, IEEE/ACM Transactions on*, vol. 16, no. 1, pp. 77–90, Feb. 2008.

[14] L. Li and R. Liu, "Securing Cluster-Based ad hoc networks with distributed authorities," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 10, pp. 3072–3081, Oct. 2010.

[15] L. Wu and B. Landfeldt, "The problem of placing mobility anchor points in wireless mesh networks," in *Proceedings of the 6th ACM international symposium on Mobility management and wireless access*, ser. MobiWac '08. New York, NY, USA: ACM, 2008, pp. 19–27. [Online]. Available: http://doi.acm.org/10.1145/1454659.1454663

[16] X. Zhu, Y. Fang, and Y. Wang, "How to secure multi-domain wireless mesh networks," *Wirel. Netw.*, vol. 16, no. 5, pp. 1215–1222, Jul. 2010. [Online]. Available: http://dx.doi.org/10.1007/s11276-009-0198-6

[17] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," RFC 3561 (Experimental), Internet Engineering Task Force, Jul. 2003. [Online]. Available: http://www.ietf.org/rfc/rfc3561.txt

[18] D. Johnson, Y. Hu, and D. Maltz, "The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4," RFC 4728 (Experimental), Internet Engineering Task Force, Feb. 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4728.txt

[19] F. Li and Y. Wang, "Routing in vehicular ad hoc networks: A survey," *Vehicular Technology Magazine, IEEE*, vol. 2, no. 2, pp. 12 –22, june 2007.

[20] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in sparse vehicular ad hoc wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 8, pp. 1538 –1556, oct. 2007.

[21] T. Taleb, E. Sakhaee, A. Jamalipour, K. Hashimoto, N. Kato, and Y. Nemoto, "A stable routing protocol to support its services in vanet networks," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 6, pp. 3337 –3347, nov. 2007.

[22] J. Nzouonta, N. Rajgure, G. Wang, and C. Borcea, "Vanet routing on city roads using real-time vehicular traffic information," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 7, pp. 3609 –3626, sept. 2009.

[23] C.-C. Ooi and N. Fisal, "Implementation of geocast-enhanced aodv-bis routing protocol in manet," in *TENCON 2004. 2004 IEEE Region 10 Conference*, vol. B, nov. 2004, pp. 660 – 663 Vol. 2.

[24] Y. Xue and B. Li, "A location-aided power-aware routing protocol in mobile ad hoc networks," in *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, vol. 5, 2001, pp. 2837 –2841 vol.5.

[25] S. Grafling, P. Mahonen, and J. Riihijarvi, "Performance evaluation of IEEE 1609 WAVE and IEEE 802.11p for vehicular communications," in *Ubiquitous and Future Networks (ICUFN), 2010 Second International Conference on*, Jun. 2010, pp. 344–348.

[26] J. Kenney, "Dedicated Short-Range communications (DSRC) standards in the united states," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162–1182, Jul. 2011.

[27] J. Zhang, "A survey on trust management for VANETs," in *Advanced Information Networking and Applications (AINA), 2011 IEEE International Conference on*, Mar. 2011, pp. 105–112.

[28] M. Almulla, H. Yahyaoui, and I. Kamal, "Trustpert: a reputation model for collaboration in manets using fuzzy expert systems," in *Proceedings of the Second Kuwait Conference on e-Services and e-Systems*, ser. KCESS '11. New York, NY, USA: ACM, 2011, pp. 3:1–3:7. [Online]. Available: http://doi.acm.org/10.1145/2107556.2107559

[29] M. Raya and J. Hubaux, "The security of vehicular ad hoc networks," in *Proceedings of the 3rd ACM workshop on Security of ad hoc and sensor networks*, ser. SASN '05. New York, NY, USA: ACM, 2005, pp. 11–21. [Online]. Available: http://doi.acm.org/10.1145/1102219.1102223

[30] H.-C. Hsiao, A. Studer, R. Dubey, E. Shi, and A. Perrig, "Efficient and secure threshold-based event validation for vanets," in *Proceedings of the fourth ACM conference on Wireless network security*, ser. WiSec '11. New York, NY, USA: ACM, 2011, pp. 163–174. [Online]. Available: http://doi.acm.org/10.1145/1998412.1998440

[31] X. Zhuo, J. Hao, D. Liu, and Y. Dai, "Removal of misbehaving insiders in anonymous vanets," in *Proceedings of the 12th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, ser. MSWiM '09. New York, NY, USA: ACM, 2009, pp. 106–115. [Online]. Available: http://doi.acm.org/10.1145/1641804.1641824

[32] J. Mittag, F. Thomas, J. Härri, and H. Hartenstein, "A comparison of single- and multi-hop beaconing in vanets," in *Proceedings of the sixth ACM international workshop on VehiculAr InterNETworking*, ser. VANET '09. New York, NY, USA: ACM, 2009, pp. 69–78. [Online]. Available: http://doi.acm.org/10.1145/1614269.1614282

[33] Sun Microsystems, "The J2EE 1.4 tutorial," p. 1134, 2005. [Online]. Available: http://docs.oracle.com/javaee/1.4/tutorial/doc/J2EETutorial.pdf

[34] G. Safdar and M. McLoone, "Randomly shifted certification authority authentication protocol for MANETs," in *Mobile and Wireless Communications Summit, 2007. 16th IST*, Jul. 2007, pp. 1–5.

[35] C. Chang and H. Tsai, "An anonymous and Self-Verified mobile authentication with authenticated key agreement for Large-Scale wireless networks," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3346–3353, Nov. 2010.

[36] K. Zeng, K. Govindan, and P. Mohapatra, "Non-cryptographic authentication and identification in wireless networks [Security and privacy in emerging wireless networks]," *Wireless Communications, IEEE*, vol. 17, no. 5, pp. 56–62, Oct. 2010.

[37] H. Wen, P. Ho, C. Qi, and G. Gong, "Physical layer assisted authentication for distributed ad hoc wireless sensor networks," *Information Security, IET*, vol. 4, no. 4, pp. 390–396, Dec. 2010.

[38] C.-I. Fan, P.-H. Ho, and R.-H. Hsu, "Provably secure nested one-time secret mechanisms for fast mutual authentication and key exchange in mobile communications," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 3, pp. 996 –1009, june 2010.

[39] S.-F. Hsu and Y.-B. Lin, "A key caching mechanism for reducing wimax authentication cost in handoff," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 8, pp. 4507 –4513, oct. 2009.

[40] K.-F. Hwang and C.-C. Chang, "A self-encryption mechanism for authentication of roaming and teleconference services," *Wireless Communications, IEEE Transactions on*, vol. 2, no. 2, pp. 400 – 407, mar 2003.

[41] H.-M. Sun and M.-C. Leu, "An efficient authentication scheme for access control in mobile pay-tv systems," *Multimedia, IEEE Transactions on*, vol. 11, no. 5, pp. 947 –959, aug. 2009.

[42] R. Housley, W. Polk, W. Ford, and D. Solo, "Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile," RFC 3280 (Proposed Standard), Internet Engineering Task Force, Apr. 2002, obsoleted by RFC 5280, updated by RFCs 4325, 4630. [Online]. Available: http://www.ietf.org/rfc/rfc3280.txt

[43] L. Harn and J. Ren, "Generalized digital certificate for user authentication and key establishment for secure communications," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 7, pp. 2372–2379, Jul. 2011.

[44] E. Holohan and M. Schukat, "Authentication using virtual certificate authorities: A new security paradigm for wireless sensor networks," in *Network Computing and Applications (NCA), 2010 9th IEEE International Symposium on*, Jul. 2010, pp. 92–99.

[45] A. Roy-Chowdhury and J. Baras, "A lightweight Certificate-Based source authentication protocol for group communications in hybrid Wireless/Satellite networks," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, Dec. 2008, pp. 1–6.

[46] J. J. Haas, Y. Hu, and K. P. Laberteaux, "Design and analysis of a lightweight certificate revocation mechanism for VANET," in *Proceedings of the sixth ACM international workshop on VehiculAr InterNETworking*, ser. VANET '09. New York, NY, USA: ACM, 2009, pp. 89–98. [Online]. Available: http://doi.acm.org/10.1145/1614269.1614285

[47] J. Forne, J. Muoz, O. Esparza, and F. Hinarejos, "Certificate status validation in mobile ad hoc networks," *Wireless Communications, IEEE*, vol. 16, no. 1, pp. 55–62, Feb. 2009.

[48] M. Nowatkowski and H. Owen, "Scalable certificate revocation list distribution in vehicular ad hoc networks," in *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, Dec. 2010, pp. 54–58.

[49] P. Wohlmacher, "Digital certificates: a survey of revocation methods," in *Proceedings of the 2000 ACM workshops on Multimedia*, ser. MULTIMEDIA '00. New York, NY, USA: ACM, 2000, pp. 111–114. [Online]. Available: http://doi.acm.org/10.1145/357744.357892

[50] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 4, pp. 491–502, Apr. 2005.

[51] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 12, pp. 1624–1637, Dec. 2005.

[52] L. Cao, H. Zhang, Y. Zhao, D. Luo, and C. Zhang, "Combined mining: Discovering informative knowledge in complex data," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 3, pp. 699–712, Jun. 2011.

[53] R. Wolff, K. Bhaduri, and H. Kargupta, "A generic local algorithm for mining data streams in large distributed systems," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 4, pp. 465–478, Apr. 2009.

[54] J. Liu and V. Issarny, "Enhanced reputation mechanism for mobile ad hoc networks," 2004, pp. 48–62.

[55] J. J. Haas, Y.-C. Hu, and K. P. Laberteaux, "Design and analysis of a lightweight certificate revocation mechanism for vanet," in *Proceedings of the sixth ACM international workshop on VehiculAr InterNETworking*, ser. VANET '09. New York, NY, USA: ACM, 2009, pp. 89–98. [Online]. Available: http://doi.acm.org/10.1145/1614269.1614285

[56] M. Raya, P. Papadimitratos, I. Aad, D. Jungels, and J.-P. Hubaux, "Eviction of misbehaving and faulty nodes in vehicular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 8, pp. 1557 –1568, oct. 2007.

[57] M. Gerlach, "Trust for vehicular applications," in *Autonomous Decentralized Systems, 2007. ISADS '07. Eighth International Symposium on*, march 2007, pp. 295 –304.

[58] M. Raya, P. Papadimitratos, V. Gligor, and J.-P. Hubaux, "On data-centric trust establishment in ephemeral ad hoc networks," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, april 2008, pp. 1238 –1246.

[59] P. Golle, D. Greene, and J. Staddon, "Detecting and correcting malicious data in vanets," in *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, ser. VANET '04. New York, NY, USA: ACM, 2004, pp. 29–37. [Online]. Available: http://doi.acm.org/10.1145/1023875.1023881

[60] F. Dotzer, L. Fischer, and P. Magiera, "Vars: a vehicle ad-hoc network reputation system," in *World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a*, june 2005, pp. 454 – 456.

[61] A. Patwardhan, A. Joshi, T. Finin, and Y. Yesha, "A data intensive reputation management scheme for vehicular ad hoc networks," in *Mobile and Ubiquitous Systems - Workshops, 2006. 3rd Annual International Conference on*, july 2006, pp. 1 –8.

[62] Y. Ren and A. Boukerche, "An efficient trust-based reputation protocol for wireless and mobile ad hoc networks: Proof and correctness," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 30 2008-dec. 4 2008, pp. 1 –5.

[63] B. Kovesi, J. Boucher, and S. Saoudi, "Stochastic k-means algorithm for vector quantization," *Pattern Recognition Letters*, vol. 22, no. 6-7, pp. 603–610, May 2001. [Online]. Available: http://dl.acm.org/citation.cfm?id=376819

[64] B. Yi, H. Qiao, F. Yang, and C. Xu, "An improved initialization center algorithm for K-Means clustering," in *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on*, Dec. 2010, pp. 1–4.

[65] A. Barakbah and Y. Kiyoki, "A pillar algorithm for k-means optimization by distance maximization for initial centroid designation," in *Computational Intelligence and Data Mining, 2009. CIDM '09. IEEE Symposium on*, Apr. 2009, pp. 61–68.

[66] R. Singh and M. Bhatia, "Data clustering with modified k-means algorithm," in *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*, Jun. 2011, pp. 717–721.

[67] Y. Sun, R. Lu, X. Lin, X. Shen, and J. Su, "A secure and efficient revocation scheme for anonymous vehicular communications," in *Communications (ICC), 2010 IEEE International Conference on*, May 2010, pp. 1–6.

[68] The Network Simulator project team, "ns-2," http://www.isi.edu/nsnam/ns/. [Online]. Available: http://www.isi.edu/nsnam/ns/

[69] S. Buruhanudeen, M. Othman, and B. Ali, "Mobility models, broadcasting methods and factors contributing towards the efficiency of the MANET routing protocols: Overview," in *Telecommunications and Malaysia International Conference on Communications, 2007. ICT-MICC 2007. IEEE International Conference on*, May 2007, pp. 226–230.