A Survey of Security in Vehicular Networks

Antonios Stampoulis (antonios.stampoulis@yale.edu)
Zheng Chai (zheng.chai@yale.edu)

As vehicles become increasingly intelligent, it is expected that in the near future they will be equipped with radio interfaces. This will enable the formation of vehicular networks, commonly referred to as VANETs, an instance of mobile ad hoc networks with cars as the mobile nodes. Such networks will provide the necessary infrastructure for various applications that can help improve the safety and efficiency of road traffic. Their primary differences from current mobile ad hoc networks is the high mobility of the nodes and the large scale of the network. Security and privacy must be two primary concerns in the design of vehicular networks. Poorly designed VANETs that permit serious attacks on the network can jeopardize the goal of increased driving safety. Also, a VANET design that enables third parties to collect private information about drivers, for example by making tracking vehicles a possibility, will certainly be avoided by drivers. Thus the specific characteristics of VANETs result in hard to address security issues, which make the field of secure inter-vehicular communications an interesting research topic, that is still very open and active. In this paper we will present the techniques that have been proposed so far to ensure security and privacy in such networks. We will start of by providing an overview of vehicular networks (chapter 1), presenting their special characteristics (chapter 2) and potential adversaries and attacks (chapter 3). The heart of this paper lies in chapter 4, where existing proposals for various aspects of vehicular communications are presented. We summarize and provide directions for future research topics in chapter 5.

1 *Overview of Vehicular Networks*

1.1 System Architecture

Every proposed architecture for potential vehicular networks requires vehicles to be equipped with some sort of radio interface that enables short-range wireless ad hoc networks to be formed. Most proposed solutions operate in a licensed frequency band of about 75MHz in the 5.8/5.9GHz band. The FCC has allocated such a frequency band in the US specifically for the purpose of vehicular networks. Similar bands exist in Japan and Europe. The MAC layer in these proposals is either a modified version of 802.11 WLAN or the 3G protocol extended for decentralized access. Since the current 802.11 protocol is not suitable for VANETs due to the high mobility and highly dynamic topology, a special version of it, called 802.11p is being developed by the IEEE. Also, the current 3G protocol is designed for centralized cellular networks, but in VANETs centralized infrastructure is not always present. There have been efforts to augment it with TDMA- and CDMA-based MAC protocols for decentralized access [1].

Most proposals also require that the vehicles be equipped with hardware that permits fairly detailed position information, so they are aware of their location. The standard choice for such hardware is a GPS or DGPS receiver, which has the added benefit of clock synchronization. The requirement of GPS receivers is not unrealistic, since such systems are already increasingly available in high-end vehicles. Still, GPS has its shortcomings, like not being able to provide position information inside tunnels. Other solutions for determining location used in wireless sensor networks are not applicable in the case of VANETs, since they require a lot of back-and-forth communication between nodes. The highly dynamic topology and strict real-time deadlines of VANETs prohibit such protocols.

Last, most proposals make the assumption that some sort of fixed road-side infrastructure will be in-place once such networks become a reality. Such infrastructure can help in various applications like collision avoidance, large-scale traffic scheduling and trip planning, and also in infotainment applications (like providing access to the Internet). The number and distribution of base stations required varies greatly between proposals: some proposals require base stations distributed equally throughout the whole road network, others only in intersections, and others at region borders. Though it is safe to assume that infrastructure exists to some extent and vehicles have access to it intermittently, it is unrealistic to require that vehicles always have wireless access to road-side base stations.

1.2 Potential Applications

There are many applications for vehicular networks. Just name a few important ones:

Collision Avoidance – About 21,000 of the 43,000 deaths that occur each year on U.S. highways result form vehicles leaving the road or travelling unsafely through intersections. Data transmitted from a roadside base station to a vehicle could warn a driver that it's not safe to enter an intersection. Communication between vehicles and between vehicles and the roadsides can save many lives and prevent injuries. Some of the worst traffic accidents involves many vehicles rear-ending each other after a single accident at the front of the line suddenly halts traffic. In this application, if a vehicle reduces its speed significantly, it will broadcast its location to its neighbour vehicles. And other receivers will try to relay the message further. And the vehicles behind the vehicle in question will emit some kind of alarm to its drivers and other drivers behind. In this way, more drivers far behind will get an alarm signal before they see the accident.

Cooperative Driving – Like violation warning, turn conflict warning, curve warning, lane merging warning etc. These services may greatly reduce the life-endangering accidents. In fact, many of the accidents come from the lack of cooperation between drivers. Given more information about the possible conflicts, we can prevent many accidents.

Traffic Optimization – Traffic delays continue to increase, wasting more than a 40-hour workweek for peak-time travelers. A significant reduction in these numbers could be achieved through vehicular networks. Vehicles could serve as data collectors and transmit the traffic condition information for the vehicular network. And transportation agencies could utilize these information to actively relieve traffic congestion. To be more specifically, in this application, vehicles could detect if the number of neighboring vehicles is too many and their avenges speed is too slow, and then relay this information to vehicles approaching the location. To make it work better, the information can be relayed by vehicles travelling in the other direction so that it may be propagated faster to the vehicles toward the congestion location. In this way, the vehicles approaching the congestion location will have enough time to choose alternate routes. Vehicles can also collect the data about weather/road surface, construction zones, highway rail intersection, emergency vehicle signal preemption and relay them to other vehicles.

Payment Services – Like toll collection. It's very convenient and desirable to pass a toll collection without having to decelerate your car, waiting in line, looking for some coins and something like that.

Location-based Services – Like finding the closest fuel station, restaurant, lodge etc. In fact, these kinds of services are not specific to the vehicular networks. Many GPS systems have such kinds of services already.

Some of these applications are safety-related, such as collision avoidance and cooperative driving. These applications are life-critical so the security of this category is mandatory. We should guarantee the proper operation of these applications even in the presence of attackers. Other applications are less safety-related or less specific to the vehicular networks. But they still need some extend of security, more or less.

2 *Characteristics of Vehicular Networks Related to Security*

In this chapter we will first investigate the requirements that vehicular networks must address in order to be deemed secure and suitable for widespread deployment, as well as the extra challenges that the very nature of VANETs raise. In the next part, we will present some of the characteristics of VANETs that will help in efficiently addressing those issues. Note that in this part we primarily focus on security issues for safety-related applications. Non-safety related applications have less strict requirements, and can reuse solutions for safety-related ones.

2.1 Security Requirements

One of the most important requirements for vehicular networks is authentication. It is crucial that transmitted messages originate from actual vehicles and not from virtual nodes. Otherwise, Sybil attacks are permitted, where an adversary can transmit an arbitrary number of safety-related packets from non-existent nodes, spreading false information to the vehicular network. An adversary might use such an attack to falsely identify a road as congested to vehicles following it, therefore diverting the traffic when it wouldn't be necessary. Thus, a mechanism for authenticating safety-related messages must be in place. This mechanism could also help prevent replay attacks, by including an authenticated timestamp in the message.

This raises some privacy concerns: the scheme used to authenticate messages can potentially be exploited in order to track vehicles' movements, if the authentication method reveals the permanent identities of vehicles. Since permitting third parties to track drivers' movements is a severe violation of driver privacy. A certain degree of anonymity will therefore always be required by users of such networks, and the authentication model used should ensure that this anonymity is maintained.

Still, there are cases where anonymity should not be maintained, and these are mostly liability-related cases. For example, should an accident happen, the information gathered through vehicular communications could very well provide detailed and irrefutable information to help legal investigations. This implies that vehicles are equipped with some sort of black-box-like device that records all communications. Also, it is required that each vehicle can be irrefutably identified by the authenticated messages it sent if there is reason to do so. Thus the property of non-repudiation is required, where a node sending a message cannot deny that the message originated from it. The privacy concerns mentioned above are aggravated with this requirement, because its potential exploitation by adversaries could lead to detailed tracking of vehicles. The identification of a vehicle from the authenticated messages it has transmitted should be something that only appropriate authorities are able to do. The issue of trust of these authorities is also raised, so it would be preferable if multiple authorities would have to cooperate before identification of vehicles is disclosed.

Another crucial requirement for safety-related applications is that the information sent from vehicles about their state must reflect their actual physical state. Should a vehicle be able to send false data to its neighbors, drivers could be influenced to take wrong decisions, compromising safety instead of improving it. Malicious users could for example send notification of abrupt deceleration, forcing vehicles following them to break, potentially leading to collisions. Mechanisms to ensure the accuracy of transmitted (and received) must thus be devised.

Last, most safety-related applications require high availability and strict message delivery deadlines to be met in order for them to be effective. Adversaries will always be able to reduce availability by doing denial-of-service attacks, so alternative communication means to further support availability might be required. Also, given the time sensitivity of most safety-related applications like collision avoidance, the design of the system should be such that real-time guarantees can be met.

2.2 Challenges

As is evident from the previous section, addressing the trade-off between authentication and non-repudiation versus privacy is an important challenge that a secure method of communication for VANETs faces. Time sensitivity is an added challenge, because it prohibits the use of security protocols that have high overhead or rely on multiple stages of full-duplex communication between nodes.

This second kind of protocols is also prohibited by the inherent nature of VANETs. Vehicles move at a fast rate, moving in and out of the reception area of other vehicles participating in the network. We can expect that a pair of vehicles can communicate for a limited amount of time. It is also expected that communication between nodes that have never interacted before and will never interact again will be the norm. Thus vehicular networks are very different from existing mobile ad-hoc networks where communication between pairs of nodes can be of relatively long duration. These characteristics imply that another class of protocols cannot be ruled effective in the vehicular networking setting: protocols requiring voting or consensus procedures and reputation-based schemes.

Another challenge is the sheer scale of the network: assuming worldwide and total adoption, a vehicular network would have close to one billion nodes! This rules out protocols that require pre-stored information about participating nodes or massive distribution of aggregated data to all mobile nodes (for example, distribution of certificate revocation lists is impossible). Such a vehicular network would also be a heterogeneous network at large, as different countries will have different security and privacy policies and differing availability of infrastructure; also different manufacturers will use different implementations. Coordination in such a network is difficult to be achieved, in many cases where it is needed; in order, for example, to distribute certified public and private keys to participating vehicles.

Other challenges that the implementation of secure inter-vehicle communications faces are the opposing incentives of participating parties. For example, law enforcement agencies would like to have total access to the transmitted information (e.g. to immediately fine violators of speed limits), while drivers would surely reject that. Also, the unavoidably slow rate of deployment poses the challenge that initially only a small number of vehicles will be part of such networks, making safety or other applications less effective. Last, such a network will have a very low tolerance for errors if we take under consideration its scale and the life-and-death importance of decisions based on information it provides.

2.3 Mitigating Characteristics

Although it is evident from the above that security in vehicular networks faces a multitude of serious challenges, such networks possess special characteristics that will be of help in addressing them.

First of all, mobility in VANETs can sometimes be a beneficial factor that helps to meet the efficiency goals. This is true in other mobile ad-hoc networks as well, but can be of even more help in VANETs.

For example, vehicles traveling in one direction can inform vehicles traveling in the opposite direction for safety conditions on the road ahead of them, in a very efficient manner. This is because a single broadcast (from vehicles traveling in one direction to vehicles in the opposite direction) can be used instead of a series of broadcasts between vehicles traveling in the same direction as the condition exists. Also, we must note that VANETs have very specialized mobility characteristics and nodes' positions in such networks are constrained within well-specified limits: roads.

Also, unlike other wireless networks like cellular networks, there is no need for confidentiality in vehicular networks. This is because messages are to be broadcast and processed by all vehicles or base stations they reach, and do not contain secrets that are to be processed by specific destinations – at least for safety-related applications. Messages therefore do not have to be encrypted prior to transmission or decrypted on reception, something which helps in meeting real-time guarantees.

Another characteristic that distinguishes vehicular networks from other mobile ad-hoc networks is that they are not as limited in power consumption. Cars have ample supply of power compared to battery-powered cell phones or sensors. Thus the communication protocols do not have to be as power-efficient; also CPU power consumption is not an issue, so we can expect significant computing power to be available. The protocols designed can therefore use frequent and complex cryptographic operations, in opposition with other mesh networks where minimization of such operations is desired.

Protocols enabling secure VANETs can also benefit from the fact that all vehicles have to be registered in a central authority, which makes possible, for example, the assignment of unique identities to vehicles (and to vehicles only). Also, vehicles undergo regularly scheduled inspections, something which permits sanity checks to be run against the components of the vehicular networking system of each car. For example, firmware can be checked for integrity and updated to the manufacturer's latest version, or malfunctioning sensors that provide false data (which may deem the vehicle an adversary) can be replaced. We can also expect that the fraction of drivers who will tinker with the internals of such systems to be very limited, as most people feel very uncomfortable modifying digital control-related subsystems of their cars. These characteristics imply that the majority of nodes will be honest, and nodes that are not will eventually be sanitized.

Another special characteristic of VANETs is that vehicles can leverage the additional input they get from the driver's response to information provided by the networking subsystem. In many situations a human driver can do a better assessment of a potential situation than what the networking subsystem is able to. As an example, an adversary might spread false information on the network indicating that he came to an abrupt stop, in order to force cars following him to decelerate or come to a halt. A human driver can reliably check whether this information is correct or not by looking at his brake lamp and at whether or not his vehicle is actually coming to a halt. The security subsystem can then interpret the driver's handling of the situation in order to check whether or not the information provided through the network was legitimate or not, and adapt its algorithms accordingly.

Last, the existing law enforcement mechanisms are likely to be extended to cover malicious behavior in vehicular networks that compromises drivers' safety, should these networks become widespread enough. This can serve as a serious deterrent to attackers of the VANET, and is a factor that is not available in other forms of wireless ad-hoc networks.

3 *Possible Attacks*

3.1 Adversaries

To better understand the different kinds of possible attacks, we need to classify our adversaries first. Because the nature and the resources of the adversary will largely determine the scope of the defenses needed to secure a vehicular network.

Greedy Drivers – It's reasonable to hope that most drivers in the system could be trusted that they will follow the protocols specified by the application, in other words, they are good drivers. But we can't deny some of the drivers will attempt to maximize their gains, regardless of the cost to the whole system, that is, they are selfish drivers. The similar situation happens in Internet, or more generally, in every social model. So the good news is we could borrow some idea from other models. For example, in our congestion avoidance system, a greedy driver might try to convince his neighbors that there is congestion ahead, and if his neighbors choose other routes, our greedy driver will get a terrific driving condition.

Snoops – This category of adversary includes everyone who attempts to collect information about you. While data mining is acceptable over aggregate data, but for identifying information for an individual, that raises serious privacy concerns and is not acceptable.

Pranksters – Pranksters include bored teenagers probing for vulnerabilities. For example, a prankster targeting a collision-avoidance might sit by the road and convince one vehicle to slow down while convincing the vehicle behind to speed up. A prankster could also abuse the security vulnerability to DoS attacks to disable applications or prevent critical information from reaching another vehicle.

Industrial Insiders – Attacks from insiders can be very harmful, and the extent to which vehicular networks are vulnerable will depend on other security design decisions. For example, if mechanics can update the firmware of a vehicle, they also have an opportunity to load malicious firmware. If we allow vehicle manufacturers to distribute keys, then a insider at one manufacturer could create keys that would be accepted by all other vehicles.

Malicious Attackers – This kind of attackers deliberately attempt to cause harm via the applications on the vehicular network. Normally, these attackers have specific targets, and they have access to more resources than other attackers. They are more professional. For example, a terrorist might manipulate the deceleration warning system to create gridlock before detonating a bomb. In general, although such kind of attackers will be less than other kinds of attackers, they are probably the most important concern for our security system.

From another perspective of view, we also can classify attackers based on some properties. We can classify attackers according to their membership of the network: if they are authenticated members or not. Obviously an authenticated attacker has more ability than unauthenticated attackers. We can also classify attackers according to their intentions: if they are malicious or rational. Unlike malicious attackers, rational ones seek personal profit and is more predictable in terms of the attack means and targets. We can also classify attackers according to their activeness or passiveness. Active ones not only listen, but also send.

3.2 Attacks

There are so many different kinds of attacks that we cannot enumerate every possible one. But like we try to classify our adversaries based on their nature, we can also try to classify attacks based on their nature, target, impact and scope to see if our applications are robust enough for these possible attacks.The most obvious attack we can imagine may be an adversary send some false information and try to convince other drivers and the system. For example, a greedy driver might pretend to be an emergency vehicle to speed up. An attacker can also send wrong information in the network to affect the behavior of other drivers. Or an attacker can send a message containing the wrong information about his identity, location, etc. An attacker can also report a non-exist jamming ahead to gain a better driving condition. In fact, this kind of attack is used nearly by all kinds of adversaries with to different extend and with different sophistication. And this kind of attack has a long history in other wireless networks. The only difference is in vehicular network, the consequence of the attack will be more serious. The most famous ones include delaying a message, replaying an earlier message or change some fields within a message. Another scenario in vehicular network is a adversary may collect "vote"s from other vehicles in normal driving, but then change the location and timestamp information in these "vote" messages and replay them at the same time to make the system believe a heavy congestion happened there. Attackers may also use this attack to alter their own position, speed, direction, etc. in order to escape liability in the case of an accident. An attacker may also pretends to be another vehicle by using false identities and do some evil he wants. Defending against this kind of attack in a vehicular network is very hard, because the traditional way of using strong identities along with cryptographic authentication may conflict with the need to preserve the privacy of participants in the network. But on the other hand, the good news for us is the computation resources to achieve this is not a problem under vehicular network scenario.

Unlike sending some modified messages, which may be not very easy to achieve if we deploy some authentication mechanism, an adversary can attack the system by sending authenticated messages, but in a improper manner. The most famous attack of this class is DoS (Denial of Service). Dos not only happens in other wireless networks, but also in other wired networks. In this scenario, the attacker try to overwhelm a vehicle's network/CPU resources or jam the whole channel of the communication network so that all the critical information cannot be delivered. The main purpose of this attack is to make the system useless, but for the real-time nature of the vehicular network, useless of the system for a short period of time may cause great danger to the driver if he is too depend on the information from the application, which is possible since human being tends to depend on technologies. Several cases have happened in GPS system. Some people are so dependent on the GPS system that they ignored the real road conditions. We can imagine a malicious adversary could provoke an accident and then use DoS to prevent the vehicles behind from deceleration. The techniques to achieve Dos include channel jamming and aggressive injection of dummy messages.

Instead of sending falsified messages and sending too many messages to overwhelm the system, sometimes an adversary can attack the system by selectively dropping the important ones. The weakness of this kind of attack is in many cases, there are more than only one or two vehicles in a certain area. Since the attacker is not the only one to relay the critical messages, this kind of attack may not be as harmful as DoS or some other kinds of attacks.

We can classify attacks according to their scope. The scope of an attack is limited if the victims consist of a small area of nodes. We call an attack affecting a larger area of nodes an extended attack. Of course an extended attack is more harmful to the whole security system so we try to keep a local attack from growing into an extended attack through information propagation. But note that for the case of limited scope attack, the victim area may not be the same area as the malicious nodes resides in. It may be another area far away from the malicious node. One may wonder how could that happen. The situation is that there exists powerful adversaries who can communicate over long distances. Such kind of adversaries have more choice on whom to convince of false data. The good news for us is although these adversaries may have more success in convincing local neighbours since there is less conflicting data nearby, but the proximity is not easy to get. The adversary has to send many vehicles to really achieve the proximity. That's not very practical. So normally we don't need to worry too much about local attack. It's true that an adversary may have more remote targets, but the correct data received from good neighbours nearby will make these attacks fail.

The result of an attack normally has three possibilities: success, detected but not corrected, detected and corrected. The most favorable situation for the adversary is an victim vehicle are isolated and surrounded by malicious nodes. This is the most rare case, and the victim don't have any choice but to accept any false messages. Basically, the only thing we can count on here is the driver itself. Our victim driver should acts like there's no vehicular network at all, which is the case nowadays. The less favorable situation for the adversary is although it's detected by some nodes to be malicious, but due to insufficient information, the good node may not be able to correct it. The least favorable situation for the adversary is not only it's detected, but also been corrected because the good node has so many honest nodes around.

4 *Proposed Solutions*

4.1 Security Hardware

Hardware is an essential part of the whole security system. Vehicle alone is just the combination of many mechanical parts. It doesn't have any computation ability. Vehicles equipped with many electronic components have some extend of computation power. But with only general electronics parts, it's hard to achieve security in a vehicular network. Just see the situation in PCs and Internet, security is an important issue. Since the vehicular network consists of nodes (vehicles), trying to keep the integrity of every node is the basis to achieve security in the whole system.

Like every other security system, some kind of log mechanism is a must. Otherwise, there's no way to track down what happened in accident reconstruction. Similar to the idea of black box in an airplane, EDR (Event Data Recorder) is introduced to our security system. Basically, EDR is a non-volatile tamper-proof storage hardware system used to record all the emergency-related information received through the vanet, including position data, speed data, acceleration data, time etc. So later if some investigation is done then these messages can be extracted, and help in the investigations. In fact, EDR is not far from us, some trucks have already installed EDR.

Only storage is not enough, there must be some kind of device generating the data to be stored in EDR and also generating and receiving encrypted messages. So we introduced the concept of TPD (Tamper-Proof Device). TPD provides the ability to verifying and signing messages. The difference between a TPD and a general CPU is that TPD not only provides the ability for processing, but also have some kind of hardware protection so it's not easy to be hacked by someone who is not authorized to do so. The basic idea of a TPD is rather simple, it contains a set of sensors to detect hardware tampering. If someone try to open a TPD by brute-force, all the stored keys in it will be erased to prevent them from being compromised. It's kind of like Cryptex in *The Da Vinci Code*. In many cases, EDR is a part of the TPD. The TPD should be as independent as possible. Normally, a TPD has its own battery which will be recharged from vehicle if necessary. A TPD will also has its own clock which can be synchronized with trusted base stations. The main problem with TPD is their cost. The current commercial products cost several thousand dollars, which is a great fraction of the whole vehicle cost. To solve this problem, we can introduce some kind of basic version TPD which provides less function at lower cost. We can also anticipate the cost of a TPD will decrease significantly with the increasing of the demand.

There's another way to solve the cost problem of TPD. We can utilize TPM (Trusted Platform Module) instead. Typically, a TPM can resist software attacks but not hardware tampering. A TPM typically provides secure storage for cryptographic keys, implementation of cryptographic primitives, and a random number generator. It also can be used in order to certify whether a signed binary has been changed or not, and to only execute it if it hasn't been. This will be useful in only permitting manufacturer-sanctioned software to be run. TPMs have been used in nowadays notebooks and cost only a few tens of dollars. The problem with TPM is obvious, it's not as tamper proof as TPD. So the most possible outcome may be a compromise between TPD and TPM.

4.2 Authentication, Non-repudiation and Privacy

The solution that is most widely considered to be the best to handle the authentication of legitimate messages requirement of VANETs is based on digitally signing every message before transmission and verifying the signature before taking a message under consideration. Digital signatures assume some cryptosystem based on public-private key pairs and a one-way cryptographic function from message space to hash-codes of specific size. Before transmission, the hash function is applied to the message, and the resulting hash is encrypted using the private key. This resulting digital signature is then transmitted along with the message, and it can be verified by the receiver for validity. The verification process involves applying the decryption algorithm using the sender's public key to the digital signature, and comparing the resulting hash with the hash produced by applying the one-way function to the received message. Only if the two hashes are equal is the message legitimate. Such a digital signature scheme addresses authenticity of messages; a third-party that doesn't have access to the origin's private key can never produce a legitimate signature for a fabricated or modified message. The process can also be enhanced in order to include some sort of timestamp in the message before signing it, having the receiver discard messages with very old timestamps, in order to protect against replay attacks. As is noted in [5], using nonces and sequence numbers instead of timestamps is not an option, because the overhead in initializing or maintaining them cannot be handled by vehicular networks due to the cursory nature of communications in them.

Due to the scale of vehicular networks, we cannot expect that each vehicle is pre-loaded with the public keys of all the other vehicles. The fact that the key pairs do not have to be static further suggests that this would not be a valid solution. Also, a model of trust like the one used in PGP, where a graph of trust is built in a decentralized manner cannot be used either for more or less the same reasons. Such a model would also not help in liability-related cases. In order for this solution to address the liability and non-repudiation requirements too, it is crucial that the public-private key pairs are issued by a central, trusted, certification authority. The certification authority can provide each vehicle with public-private key pairs as well as certificates for the public keys. Such a certificate includes the actual public key and also a digital signature, signed with the CA's private key. This signature includes both the public key and the unique identifier of the CA. Since vehicles receiving a signed message cannot validate it without the public key of its originating vehicle (which they do not necessarily possess), the public key must be sent along with the message. Of course, this is not enough – in order for the public key to be trusted, it must have been provided from the trusted certification authority. That's why each message must be accompanied by the full certificate provided by the CA. Thus the full protocol for authentication of received messages is: first extract the public key out of the certificate and check whether it is issued by the expected CA using the signature included in the certificate. Then, use this public key to validate the signature of the message. At this point, the message is deemed authentic.

To see how this scheme meets the non-repudiation requirement, we must first talk about vehicle identification. It is expected that each vehicle will be characterized by a single unique identifier, issued by either the respective government or by the manufacturer. In the first case, this identity is very much like the physical vehicle identification in the form of license plates, and is called the Electronic License Plate (ELP). We assume that this is assigned to each vehicle at the time of registration with the local government, and is changed in the case that the owner changes or that the original owner moves to another country. In the second case the identity of the vehicle is called the Electronic Chassis Number (ECN), and each vehicle is provided with one at production time. The certification authority can therefore store a mapping between the unique identity of a vehicle and the set of public keys it provided that vehicle with. When a liability case arises, the CA will thus be able to provide the public keys of associated parties to the law enforcement authorities, should these authorities have proper permission. Using these keys the origin of each related message (stored in the Event Data Recorder of each vehicle) can be validated in an irrefutable way.

Let us now consider how this approach to addressing the authentication and non-repudiation requirements rates in regards to preserving the anonymity of drivers. First of all, the key pairs provided to the vehicles by the CA should not have any relationship with their unique identities – this relationship should only be determined by means of actual disclosure of identities-key pairs matching by the CA on liability cases (as described above). Otherwise, vehicles' identities can be determined and drivers' privacy is violated. Also, equipping each vehicle with only a singleton or small static set of key pairs compromises drivers' privacy too, because a third-party can track the vehicle by tracking messages signed with the same public key. Therefore, the keys vehicles use to sign their messages must be used for a limited amount of time.

There are multiple ways to achieve that. Maybe the simplest way is the one proposed in [4], according to which vehicles must be provided with a large set of public/private keys (and the corresponding certificates) during the process of yearly security inspections. These keys are then used for signing messages, making sure that each key is only used for a brief period of time. During that period, vehicles' movement can be tracked, but if it is brief enough, this is not a privacy violation as the next key used cannot be matched against the previous one. In [5], the period where each key is used adapts to the speed of the moving vehicle, and it is analytically derived that a key use period of about one minute is safe in ensuring the anonymity of vehicles. This solution is effective, but requires a large set of keys and certificates to be stored in the vehicle (although the authors show that in typical scenarios this would be on the order of a few megabytes) and also, provides potential attackers with a large set of certified keys which they can use to launch Sybil attacks.

Other ways to achieve anonymization are proposed in [3]. One is to have a public-private key pair stored in the TPM of the vehicle to serve as the vehicle's identity, and a certificate for this provided by the vehicle's manufacturer. Then the vehicle would regularly communicate with an online CA in order to get certificates for anonymous key pairs, providing this identity and certificate. The CA should only certify a key if it trusts the vehicle manufacturer, and should not certify a new key before a previous one is close to expiration, to prevent Sybil attacks. This approach has the obvious drawback of requiring an online CA able to process very frequent certificate requests; it is not entirely sure that a scalable implementation of such a scheme is possible. Another way to achieve anonymization is to have each car authenticate at various points (for example, in toll booths), where it would be provided with a private-public key pair that would serve as its temporary ID. Then, it would communicate with anonymization services provided by frequent roadside base stations, in order to get a temporary private-public key pair with even shorter usage period (as in the above solution). This solution is effective in ensuring anonymity and preventing Sybil attacks, but it requires significant infrastructure to be present (frequent authentication points and very frequent base stations). Also, ensuring that the non-repudiation property is held is not trivial. Gathering and matching data about associated parties in a liability case both from authentication points and anonymization services at base stations will be required. This would also mean that such data for all the vehicles with which authentication points and base stations interacted must be kept. This is a significant space overhead and might deem this solution not scalable enough.

Here we should note that key pairs and certificates should be stored inside the TPD, to prevent malicious users from gaining access to them. The process of signing messages prior to transmission and signature validation of received messages must also be handled by the TPD, to minimize the harm that compromised higher-level firmware could cause to the VANET. In [5] it is also shown that TPDs can be used to help handle revocation of certificates, in the case where a vehicle's user is identified as malicious and should therefore be removed from the network. Considering the above, an outsider is limited to jamming attacks, so revocation of certificates would be an effective security measure. Distributing certificate revocation lists to all vehicles is not an option, as noted in chapter 2. The ways proposed in [5] to handle certificate revocation are the following. In the first, a message is sent from the CA to the malicious vehicle's TPD that forces it to delete all its keys. Unable to sign outgoing messages, the node cannot send legitimate messages in the VANET. In the second, the CA wants to invalidate only specific key pairs of a vehicle, and distributes compressed certificate revocation lists using probabilistic Bloom filters to force the vehicle to revoke those keys and notify neighbors of the revocation. In the last, neighbors of malicious vehicles identify them as such by accumulating accusations, and report them to the CA.

A potential shortcoming of the authentication scheme we described here is that the overhead of adding a digital signature and a certificate to each message sent might be significant. To alleviate this problem, one might opt to use a more space-efficient cryptosystem like ECC instead of RSA. Also, the scaling of such a scheme is a concern, since the CPU of a vehicle will have to verify digital signatures and certificates at a fast pace with hard real-time guarantees. This overhead is not trivial, but as we mentioned above, the CPUs of vehicles can be of considerable computational power, so this is not so much of a problem as it would be in, for example, a cellular network. Also, the processor can skip the verification of messages that it deems as not important or not relevant. Last, it is shown in [5] that this authentication scheme is more efficient than using symmetric encryption between pairs of vehicles when considering the large scale of the network. In effect, the initial overhead of establishing keys between communicating vehicles amounts to more overhead than signing messages with asymmetric encryption schemes.

4.3 Authentication of Aggregated Data

In applications where flooding the network with information is required, the protocol used to forward messages from one vehicle to another can create significant overhead. This is the case for example in emergency road condition warning applications. In applications like that, it is expected that a vehicle at some point will receive a multitude of messages from vehicles closer to the point where the emergency road condition exists. This vehicle now will have to forward this information to vehicles further back the road. A simple protocol that just forwards all messages to other vehicles will be extremely inefficient, because in the general case, a lot of redundant information will be propagated. Messages related to the same road condition will be similar enough; so forwarding just one of them will contain all the information needed. Even if that is not the case, we can expect that some sort of summary or compression of related messages is possible and can greatly reduce the size of the message that is to be forwarded, while still retaining all the needed data for vehicles further back the road. Such a case would, for example, be an application for congestion avoidance, where multiple cars would transmit their position and velocity. An easy way to summarize such data would be to use less precision to describe positions as we get further away from the originating nodes (syntactic aggregation). Thus the issue of how to authenticate messages that aggregate information provided by messages originating from other nodes is raised. This is not entirely trivial because we want to be sure both of the fact that all the individual messages were authenticated, and of the fact that the message containing the aggregated data is authenticated.

In [7], the authors propose ways to address a similar issue, examining only the case where the individual messages are identical. They assume that there is some way for vehicles to dynamically form groups – we examine this issue in the next section. The issue they try to solve is how to avoid having each vehicle in a group transmit a message with essentially the same data in order to reduce the channel usage. We can directly apply their solutions to the case where a vehicle wants to avoid forwarding multiple copies of the same message, by having the vehicle delay forwarding a message it receives for a short period, in which identical messages it receives are discarded. Note that this problem is not entirely trivial – the vehicle cannot just discard the signature from the message and forward the message signed with its own signature, because in that way, the information contained in the message can be regarded as malicious: the vehicle transmitting the information is far enough from the point where the condition exists, so it shouldn't be able to tell that such a condition exists if it has not been informed by vehicles closer to it. Thus the vehicle forwarding the message must provide authentication for all the vehicles in the path from where the condition exists to itself. According to the first solution in [7], instead of transmitting multiple identical copies of a message along with the signature and certificate of their originating vehicle, one should only transmit one copy of the message, along with a concatenation of the signatures for it by all the intermediate vehicles (including itself, the forwarding node), as well as all the certificates. In that way, nodes receiving the message can verify that all the individual nodes have agreed to the message's contents by validating the signatures. This reduces the overall channel usage significantly, but raises another problem: the space overhead of including all signatures and certificates becomes non-trivial. Even as cryptosystems have become sophisticated enough so as to make this overhead small enough for practical applications, the fact that it is repeated many times (for all individual vehicles that have sent the message), makes this overhead substantial. Thus a mechanism to reduce this overhead is needed.

Such a mechanism is provided by the authors in the form of onion signatures. The idea is, instead of having the vehicle sign the message and concatenate its signature to the existing list of signatures, to have each vehicle sign the current signature – creating what they call an onion signature. The message is forwarded along with the previous signature, the current onion signature, and the list of certificates. Since both the previous and the current signatures are included, the vehicle receiving the message can check whether the forwarding was done by an authenticated vehicle. Also, it can check, by repeatedly applying the public keys from the certificates to the onion signature, whether the original message was authenticated. This mechanism will reduce the space overhead imposed by the multiple signatures considerably, because the message always includes a small constant number of them – just two signatures. This is not without compromise of course, because the computational complexity of authenticating such messages is increased. As we said before, this is not so much of a concern in the setting of VANETs. Also, if one vehicle in the chain of forwarding vehicles is malicious, and fails to provide a valid signature, the message is discarded since it cannot be authenticated anymore – because of the way onion signatures work.

Another way to deal with authentication of aggregated data is proposed in [8]. This proposal can also handle messages that are similar but not identical, and expects nodes receiving multiple messages with similar information to summarize the information in them using only syntactic aggregation. This means that the information of all the messages is retained in separate entries, but can be compressed or reduced in precision. The main idea that the authors propose is to challenge the forwarding vehicle to provide probabilistic proof that the aggregated message is authentic and not constructed. We assume that the TPD provides a transmit buffer service where applications place messages to be transmitted. The TPD signs and sends these messages after a small delay, during which the applications can append data to them. Also, it provides a trusted random number generator service. The proposed protocol works as follows. The application that aggregates the data puts the summarized message in the transmit buffer. This includes N entries, one for each incoming message that is summarized. The TPD includes a random number in the message, signs it, and is ready to transmit it. In the meantime, the application must read the random number, and include the original message of index i along with its signature, where i is determined by applying some function to the random number (scaling it to range 1 to N). In case this is successful, the message sent will include both the aggregated data and one original message, which serves as probabilistic proof of the authenticity of the data: it can be verified by other nodes that this message is authentic and the information contained in it agrees with the aggregation. In case it isn't, it means that the sender of the aggregated message was malicious, since that's the only case where the original message and its signature could not be produced. Since the transmit buffer will transmit the message anyway, the transmitted message serves as proof of malicious behavior. This solution is very interesting, in the sense that it only requires minimal overhead and is quite effective in ruling out malicious behavior, as preliminary analysis in the paper shows. Extensions to handle semantic aggregation are still to be investigated.

4.4 Group Formation and Communication

Every security system faces a same problem: efficiency. A good system design is trying to achieve security without too much overhead. But how can we make the system more efficient without sacrificing security? How can we achieve these two contradictory requirements simultaneously? The idea of having vehicles communicate through groups and not individually seems to be a good candidate. Roughly speaking, instead of communication among all the vehicles, which may cause great overhead for the whole system, we can divide the vehicles into groups, and let the communication happen mainly among group leaders. This approach of communication in groups has already been suggested in the previous section, and it would improve the efficiency in scenarios described in it. Also, it would enhance the preservation of anonymity in the resulting protocol. The questions that arise are how groups are formed, how communication happens inside groups without sacrificing security, and how groups reach consensus.

In general, group formation can be predefined or dynamic. In the first case, specific vehicles are part of specific groups. This approach is clearly too rigid and not scalable, thus not suitable for VANETs, but might be of use in some corner cases like having all public transport vehicles be part of a group. In the second case, groups are formed dynamically, presumably based on how close they are or what their driving pattern is. How group formation is done in this case in the VANET setting is not clear at this stage, as it hasn't been investigated yet in papers dedicated to this issue. It is a difficult issue, because the overhead of group formation must be very small, since vehicles are likely to be part of a group for a small amount of time. A hybrid approach is proposed in [7], where geographic-based group formation is defined. To be more specific, the map is divided into small cells. Both the map and the cell partitioning can be built into the GPS system of every vehicle. So every vehicle knows which group it belongs to at any moment based on its location. In order to keep every vehicle in touch with at least one group leader while moving, every cell overlaps with nearby cells. Vehicles in the overlap area also serve as communication bridge between the two groups. Since the location of every cell is determined, then the location of the center of every cell is also determined. The vehicle which is closest to the center will be assigned the group leader for that cell. It is assumed that vehicles periodically transmit a safety-related message containing their location, so all vehicles in the group know, more or less, the public key and position of all other vehicles in the group. Thus, the group leader is unanimously agreed upon by all vehicles in a cell without extra communication. In this way, we decrease the overhead of message transmission without causing another kind of overhead, group formation and leader election. And we can easily use geographic routing in this scheme.

Communication in groups can be done using symmetric cryptography in order to reduce both the space overhead and the computational complexity. This needs a symmetric cryptographic key, that must be shared between group members. A way to establish such a key would be to have members communicate in order to mutually decide upon one. But this approach would require serious communication overhead and it would take a lot of time to reach consensus. A better way to establish a key with small communication overhead is have the group leader choose a key and inform all members about it. The group leader distributes the group key to all the group members. And subsequent message broadcasts include only a hashed message authentication code in addition to the message, which is a very small overhead. When a new group member enters, it will receive the group key from the group leader. But this group leader may be not the same one as before. When a vehicle leaves the group, nothing needs to be done – the vehicle will join another group and use the keys defined in it. The use of secure groups protects the network from outsiders: for a vehicle to join the group, it is assumed that an authenticated safety-related message containing the location and public key of the vehicle has been sent (so the vehicle is valid). All communications inside the group are authenticated using the established symmetric key, so vehicles who have not joined the group cannot send a valid message to it. Still, the non-repudiation requirement is not met, because messages inside the group have no information that differentiates one vehicle from the other. It is probable that in the case of fully dynamic groups, the group leader can communicate with an online CA in order to establish a certified key for group communication, plus temporary unique IDs for all involved vehicles, for non-repudiation purposes. This approach is yet to be fully developed.

Group agreement is an even more demanding problem. Normal algorithms used in sensor networks are not suitable for use in VANETs, because they require a lot of back-and-forth communication between group members. We have found no existing literature on this issue with regards to VANETs. While a demanding problem, it is mitigated by the fact that the majority of vehicles in a group (assuming the group formation scheme described above) will share a similar view of their environment if they all process the various events locally. This is because most members of a group will receive similar messages from other vehicles, and will gather similar data for events happening inside the cell they're in using their sensors. Also, adversaries that want to spread false data (who therefore have a substantially different view of the environment) are expected to be a small minority, and their false data can be detected using techniques described in the next section. Thus, the majority of members of a group will inherently roughly agree on the information that is to be transmitted, therefore reaching consensus should not be as hard as the general case. A protocol leveraging these characteristics is still to be described, so this problem is an open one.

4.5 Detection and Correction of Malicious Data

In some situations the requirement of only considering messages with information that reflects accurately the actual physical environment is very strict, and ruling out false information is essential. Despite the security provided by the combination of TPDs with authenticated messages, an attacker could overcome it and manage to transmit valid messages containing false data, albeit with substantial effort. Such a case would be, for example, if the hardware sensors are tampered with, like putting the vehicle temperature sensor in cold water. To rule out such cases of false data, we should not only verify that the sender of the data is legitimate, but also that the data per se are legitimate. Therefore some proposals on detection of malicious data have been made. Please note that the sender of the data may not necessarily be malicious, but his vehicle's sensors may be broken. Still, we refer to all false data as malicious, because it can potentially cause harm, even if the sender doesn't have malicious intentions.

The first proposal focuses in verifying the position data that vehicles send, since this is probably the most important information that is shared in the VANET, and is presented in [9]. The authors of the paper define a number of sanity checks that any vehicle can perform locally on position information it receives from other vehicles. All position information received from a vehicle is stored for some time period; this is used to perform the checks, the results of which are weighted in order to form a metric of the neighbor's trust. We will briefly describe the checks here.

The first is checking whether the claimed position of a vehicle is within the acceptance range of the receiving vehicle's radio interface. Assume vehicle A broadcasts the message, and vehicle B receives it. Should the vehicle A claim that it is outside this range, the position information is classified as false, since if it was true, vehicle B would not be in position to successfully receive A's message. This check rules out cases where a malicious node claims that it is in better position to forward a message, although it is in a position quite far away.

The second is to compare subsequent beacon messages sent from A to see whether given the position in the first message, it is possible that it has reached the position in the second message, given some maximum speed limits. (Beacon messages are safety-related messages that are sent periodically, containing the position of the vehicle and potentially other information too.) This prevents interception attacks where a malicious node is sending true position information up to some point and then wants to intercept a message. In that case it would send a beacon message with a position superior to forwarding. If the difference in positions is large enough so as to be impossible with the imposed maximum possible speed limit, the position information is identified as malicious.

The third check is a simple density threshold check: because of the physical dimensions of vehicles, only so many can be in a specified area. If successive messages with position information suggest that an area is more densely occupied than is physically possible, they are discarded as malicious. The last check is a map-based one. Assuming that all nodes have access to information about street maps (for example, using a navigation system), the claimed position of a car can be checked for plausibility. For example, if a vehicle claims that is inside a house, its position information should be classified as unlikely.

A more complete proposal that handles both detection and correction of malicious data is given in [10]. The authors' primary assumption is that the simplest explanation of some inconsistency in the received information is most probably the correct one. Their idea is the following: each vehicle maintains a model of the VANET, containing information about the actual physical state of the network, based on the messages it has received. This model can be checked for consistency using checks similar to the ones described above. If a new message is received, the vehicle tries to incorporate the information contained therein in its existing model. If this renders the model inconsistent, then the vehicle searches for the minimal set of malicious vehicles and messages, that if removed from the model, would make the model valid again. These vehicles and messages are removed from the model, and the process continues. The authors show that this proposal is very effective in handling various attacks when accurate position information for neighbors can be determined through means other than communication. If this is not the case, and only a distance metric can be determined, the authors show that it is expected to be effective enough for many attacks. Still, much work needs to be done on this proposal, because many parts of it remain unspecified. For example, the minimal-set-search algorithm is only hinted at, while it is supposed to solve an intractable, in general, problem. Nevertheless, it is a proposal that will further enhance the security of VANETs, should it be fully developed and combined with the other security mechanisms we have already described.

5 *Summary*

In this paper we did a brief survey of security-related issues and proposed solutions in the setting of vehicular ad-hoc networks. We introduced the basics of the design of such networks, and the special characteristics of them. We presented what the requirements are for a secure yet efficient method of communication in this setting, where dissemination of false information affects life-critical decisions and therefore must be eliminated. We saw the potential adversaries in such a system and the attacks they might try to launch.

Then specific proposals were presented for various aspects of communication in such networks. First, we present in depth a method for communication supporting authentication and non-repudiation while still maintaining drivers' anonymity, using certified digital signatures and depending on a tamper-proof hardware module. Building on this method, ways to reduce its overhead, in applications where flooding is required are proposed, based on message aggregation or on group communication. We close the survey presenting ways on detecting and correcting malicious data, making sure that they don't interfere with safety-related applications. It is evident from our presentation that the field of security in vehicular networks is largely an open one, and a lot of issues remained to be solved, despite the fact that many interesting proposals have been made.

6 *Bibliography*

[1] Survey of Inter-Vehicle Communication, *Jun Luo and Jean-Pierre Hubaux*

[2] Security Issues in a Future Vehicular Network, *Magda El Zarki, Sharad Mehrotra, Gene Tsudik and Nalini Venkatasubramanian*

[3] Challenges in Securing Vehicular Networks, *Bryan Parno and Adrian Perrig*

[4] Securing Vehicular Communications, *Maxim Raya, Panos Papadimitratos and Jean-Pierre Hubaux*

[5] Securing Vehicular Ad Hoc Networks, *Maxim Raya and Jean-Pierre Hubaux*

[6] Attacks on Inter Vehicle Communication Systems - an Analysis, *Amer Aijaz, Bernd Bochow, Florian Doetzer, Andreas Festag, Matthias Gerlach, Rainer Kroh and Tim Leimueller*

[7] Efficient Secure Aggregation in VANETs, *Maxim Raya, Adel Aziz and Jean-Pierre Hubaux*

[8] Probabilistic Validation of Aggregated Data in Vehicular Ad-hoc Networks, *Fabio Picconi, Nishkam Ravi, Marco Gruteser and Liviu Iftode*

[9] Improved Security in Geographic Ad-hoc Routing through Autonomous Position Verification, *Tim Leinmueller, Christian Maihoefer, Elmar Schoch and Frank Kargl*

[10] Detecting and Correcting Malicious Data in VANETs, *Philippe Golle, Dan Greene and Jessica Staddon*