Universitat Rovira i Virgili

Department of Computer Engineering and Maths

Ph.D. Dissertation

Security and Privacy Issues in Some Special-Purpose Networks

Author:
Alexandre Viejo Galicia

Advisors:
Dr. Francesc Sebé Feixas and Dr. Josep Domingo Ferrer

Dissertation submitted to the Department of Computer Engineering and Maths in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science

June 2008

© Copyright 2008 by Alexandre Viejo Galicia
All Rights Reserved

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy in Computer Science.

Dr. Francesc Sebé Feixas

(Advisor)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy in Computer Science.

Dr. Josep Domingo Ferrer

(Advisor)

Approved by the University Committee on Graduate Studies:

Preface

This thesis is about providing security and privacy to new emergent applications which are based on special-purpose networks. More precisely, we study different aspects regarding security and privacy issues related to sensor networks, mobile ad hoc networks, vehicular ad hoc networks and social networks.

Sensor networks have a wide variety of applications related to event surveillance like emergency response or habitat monitoring. Two contributions providing scalable and secure transmission of sensed data are presented.

Ad hoc networks are suited for use in situations where deploying an infrastructure is not cost effective or is not possible for any other reason. When the nodes of an ad hoc network are small mobile devices (*e.g.* cell phones or PDAs), such a network is called mobile ad hoc network. If mobile nodes are embedded in cars, then that network is called vehicular ad hoc network. Different schemes providing secure and private information transmission in both types of ad hoc networks are presented.

Social networks differ from the special-purpose networks commented above in that they are not physical networks. Social networks are applications that work through classic networks. They can be defined as a community of web users where each user can publish and share information and services. A privacy-preserving resource access protocol for social networks is presented.

Contents

Preface **v**

1 Introduction **1**

1.1 Situation and objectives 1

1.1.1 Many-to-one information transmission for sensor networks 4

1.1.2 Information sharing in MANETs 6

1.1.3 Information spread in VANETs 8

1.1.4 Resource access in social networks 10

1.2 Structure of this thesis 10

2 State of the Art **13**

2.1 Many-to-one information transmission 13

2.1.1 Secure many-to-one lossy transmission 15

2.1.2 Secure many-to-one lossless transmission 17

2.2 Information sharing in MANETs 20

2.2.1 Privacy, security and incentives 20

2.3 Information spread in VANETs 24

2.3.1 A posteriori countermeasures against fake messages . . 24

2.3.2 A priori countermeasures against fake messages 25

2.4 Resource access in social networks 28

3 Efficient and secure many-to-one symbol transmission 33

3.1 Secure many-to-one communications based on GDH multisignatures 34

3.1.1 Multisignatures over Gap Diffie-Hellman groups 35

3.1.2 General assumptions 35

3.1.3 Reverse bit transmission 36

3.1.4 Security analysis 38

3.1.5 Performance analysis 40

3.1.6 Error handling 42

3.1.7 Generalization to q-ary transmission 44

3.2 Secure many-to-one communications for resource-constrained devices 46

3.2.1 General assumptions 47

3.2.2 Many-to-one q-ary transmission 47

3.2.3 Procedure to deal with corrupted messages 51

3.2.4 Security analysis 53

3.2.5 Performance analysis 56

4 Secure and private information sharing in MANETs 61

4.1 Distributed information provision 62

4.1.1 System overview 62

4.1.2 System architecture 63

4.1.3 Security analysis 71

4.1.4 Privacy analysis 73

4.2 Advertisement dissemination 74

4.2.1 System overview 75

4.2.2 Set of protocols for advertisement dissemination 76
4.2.3 Example of an advertisement dissemination 81
4.2.4 Comparison to other reward models 83
4.2.5 Security and privacy analysis 85

5 Private and trustworthy information spread in VANETs 91
5.1 Cryptographic background 92
5.1.1 Secret sharing 92
5.1.2 Threshold signatures 93
5.1.3 Privacy in secret sharing 95
5.2 Trustworthy privacy-preserving announcements in VANETs . . 96
5.2.1 Non-private protocol 96
5.2.2 Cost analysis of the non-private protocol 99
5.2.3 Group-based private protocol 103
5.2.4 Extended group-based private protocol 104
5.2.5 Semi-private protocol for sparse VANETs 106
5.2.6 Compound protocol 108
5.2.7 Simulation 111

6 Private resource access in social networks 121
6.1 Multiplicative privacy homomorphisms 122
6.2 Homomorphic access control protocol for social networks . . . 123
6.2.1 Homomorphic encryption of rational values 125
6.2.2 Simple homomorphic protocol 128
6.2.3 Anonymous homomorphic protocol 132
6.2.4 Accessing the resource 138
6.2.5 Security and privacy analysis 139
6.2.6 Simulation 142

7 Conclusions **149**

7.1 Concluding remarks 149

7.2 Results of this thesis 150

7.3 Future research 152

Our contributions **155**

Bibliography **157**

Chapter 1

Introduction

1.1 Situation and objectives

The transformation from the *Industrial Society* to the *Information Society* started somewhere between the 1970s and today. In the former, the economy is based on material goods and how to produce them. In the latter, the knowledge is the main engine that moves the progress and the development of humankind. To be useful, the knowledge must be properly managed. It means that we need effective ways to generate it, store it and process it.

The Information Society is based on the use of computers and the exchange of information between them. Two computers communicate (exchange information) through a connection (wired or wireless). Different computers connected between them form a computer network.

Computers sharing resources and cooperating between them can solve large problems which otherwise would be impossible to address by the same computers working alone [Fost02]. Networked devices can even enable the emergence of new models and scenarios which represent new opportunities and challenges. According to that, computer networks are considered an

important advance. Note that the "resource" concept includes the whole range of things that can be shared in a connected computer system: from hardware components such as printers to several sources of information: files, databases, video/audio streams, etc.

Computer networks are everywhere. The most important one is *the Internet* which is a very huge network composed of several smaller networks. The Internet allows computers to connect to other computers easily, wherever they may be across the world. This global computer network is one of the most important innovations of our time, bringing substantial benefits to economies and societies, but also driving change in the way we live and work [Euro08].

Despite the big importance of the Internet, the society can still grow. In this way, researchers continue developing new applications and their communication paradigms. Part of these new proposals are based on special-purpose networks. Some of such special networks are physically different from the classic network which consists of static full-fledged computers with permanent connections between them. Examples of these networks are the following:

- *Sensor networks.* These networks consist of resource-constrained wireless devices with sensor capabilities. This emerging technology has a wide variety of applications related to event surveillance like emergency response, habitat monitoring or defense-related networks.

- *Mobile ad hoc networks (MANETs).* Such networks are formed by mobile nodes which are connected in a self-organized way without any underlying hierarchical infrastructure. Small devices enabled with wireless communications technologies (*e.g.* cell phones or PDAs) are usually

used as nodes in MANETs. Ad hoc networks are suited for use in situations where deploying an infrastructure is not cost effective or is not possible for any other reason. One of many possible uses of MANETs is to provide crisis management services applications, such as in disaster recovery, where the entire communication infrastructure is destroyed and reestablishing communication quickly is crucial [Jhu07]. Another useful situation for MANETs is a scenario without fixed communication systems where there is the need for any kind of collaborative computing. Such situation can occur in both business and military environments.

- *Vehicular ad hoc networks (VANETs).* When the mobile nodes of a MANET are embedded in cars, such a network is called Vehicular Ad hoc Network (VANET). This kind of networks can be very useful to increase the road traffic safety and they will be deployed for real use in the forthcoming years. As a proof of that, eight important European vehicle manufacturers have founded the *CAR 2 CAR Communication Consortium* [C2cc08]. This non-profit organisation is dedicated to the objective of further increasing traffic safety and efficiency by means of inter-vehicle communications.

Sensor networks, MANETs and VANETs differ from classic networks in the hardware that they use and in the way the nodes are connected between them. Nonetheless, there are special-purpose networks which are only applications that work through classic networks. This is the case for *social networks*.

Nowadays, social networks have become an important web service [Staa05] with a broad range of applications: collaborative work, collaborative service rating, resource sharing, searching new friends, etc. They can be defined as a community of web users where each network user can publish and share

information and services (personal data, blogs and, in general, resources). Social networks have become an object of study both in computer and social sciences, with even dedicated journals and conferences.

The special-purpose networks described above provide a wide range of new services and applications. Even though they are expected to improve the society in several ways, these innovative networks and their related applications bring also security and privacy issues that must be addressed.

In this thesis, we solve some security and privacy issues related to such new applications and services. More specifically, our work focuses on:

- Secure information transmission in many-to-one scenarios with resource-constrained devices such as sensor networks.
- Secure and private information sharing in MANETs.
- Secure and private information spread in VANETs.
- Private resource access in social networks.

1.1.1 Many-to-one information transmission for sensor networks

Communications can be classified according to the number of involved senders and receivers. Single-sender paradigms are: *one-to-one* (unicast) in which a single sender transmits data to a single receiver; *one-to-all* (broadcast) in which one source sends data to all nodes of a network; and *one-to-many* (multicast) where a single source transmits to a given subset of nodes.

Efficient one-to-many (and one-to-all) communications are implemented using a tree communication model. The root of the tree is the source which

sends the data, the intermediate nodes are the routers which receive the content from their parent node and retransmit it to their child nodes (replicating it for each child), and the leaves are the receivers. This model provides scalability because the number of receivers can be increased without increasing the workload nor the bandwidth needs at the source.

At some point, one-to-many applications may require the root of the tree to collect data from all users. This situation results in many-to-one communication [Mill99]. If the number of transmitting nodes is large, the receiver may be overwhelmed by the incoming traffic. This problem is known as implosion [Quin01].

Implosion resistance is a challenging issue in the design of many-to-one communication protocols. Such protocols also follow a tree topology. In this case, the leaves are the senders; the intermediate nodes are routers that collect messages coming from their children and aggregate them into a single message that is transmitted up to their parent; finally, the root is the receiver. Scalability depends on the aggregation operation performed by intermediate routers.

In addition to their being scalable, many-to-one communications often need to be secure. Security requirements include *confidentiality* (an intruder should not be able to learn the transmitted data), *integrity* (any data alteration should be detectable by the receiver) and *authentication* (the source of the data should be verifiable by the receiver).

There is a general consensus that in scenarios where nodes are resource-constrained devices the high cost of public-key cryptography is usually not affordable. Researchers assume that in such scenarios symmetric cryptography and hash functions constitute the tools of choice to provide security.

However, public-key technology can be selectively deployed in those environments too. In [Bene05] the author argues that the RSA [Rsa78] public-key cryptosystem with a small public exponent and Rabin's [Rabi79] public-key cryptosystem have fast algorithms for encryption and digital signature verification which can be used on constrained devices. In contrast, their decryption and signature generation are slow and resource-demanding. Elliptic curve cryptosystems (ECC, [BlaB05]) provide not only lightweight encryption and signature verification, but also lightweight decryption and signature generation which make them suitable for resource-constrained devices.

Sensor networks are an example where secure many-to-one communications are required in low-cost and resource-constrained devices. Here, the sensor nodes (which may be very numerous) transmit data to a single collecting center. Security requirements arise when the networks are deployed in hostile areas.

Due to the relevance of this emerging technology, new lightweight protocols providing secure many-to-one communications should be designed.

1.1.2 Information sharing in MANETs

New-generation mobile devices (*e.g.* cell phones, PDAs ...) are enabled with wireless communications technologies which paves the way to a broad range of services based on mobile ad hoc networks.

MANETs are extremely dynamic. Thus, nodes are constantly changing their location. This can cause any pair of nodes to be temporarily unconnected. According to that, communication systems for this kind of networks should not depend on centralized authorities that need to be accessible all the time.

Information transmission between peers is a basic process in MANETs.

Such networks rely on the data forwarding service to transmit data between users. It consists of correctly relaying the received packets from node to node until they reach their final destination. A survey presented in [Djen05] states the following threads in this procedure:

- *Eavesdropping.* Malicious nodes can eavesdrop packets in transit and analyze them to obtain confidential and sensitive information. It is considered a passive attack.

- *Tampering.* Dishonest nodes can tamper with the forwarded data to get some benefits.

- *Dropping data packets.* Transmitted packets follow multi-hop routes and pass through mobile nodes. A malicious node which participates in the routing can drop all packets it gets to forward.

- *Selfish behavior.* Nodes involved in a MANET usually do not belong to a single authority and do not pursue a common goal. Therefore, nodes are not directly interested in forwarding packets for others. In consequence, there is no reason to trust nodes and assume that they will cooperate. Even though this is not an intentional attack, it is as harmful as dropping data packets.

Data eavesdropping and data tampering are commonly solved using cryptography. Cryptography-based secure communications in MANETs have been widely addressed in the literature [Djen05].

Providing incentives to relaying nodes is a way to prevent peers from dropping data packets (it can be done intentionally or as a result of a selfish behavior) [Butt00, Butt03, Sale03]. Nevertheless, provision of incentives to

nodes in a secure way for both the payer and the payee is not straightforward and must be addressed properly.

Last but not least, MANET applications may require the interaction between some devices in order to trade certain goods or services. Such interaction may imply the disclosure of some information related to the involved users. A third party collecting such information may be able to track users and obtain confidential data about their habits and whereabouts which represents a serious menace. According to that, communication protocols for MANETs must be privacy-preserving. Privacy includes anonymity and unlinkability. Anonymity refers to the requirement that a user should be able to participate in the network without revealing her identity. However, anonymity must not imply impunity for dishonest users who try to disrupt the system. Unlinkability means that different interactions between a specific user and the network communication system cannot be related to each other neither by the system nor by an external observer. Note that, if a system is anonymous but the different actions by the same user are linkable, the user's roaming pattern can be obtained from such linkage; this might suffice to infer the user's identity (the roaming pattern typically includes going home, going to a certain job location, etc.).

Due to the mobility inherent in MANETs, new applications which exploit this property must be developed. Such new schemes must provide incentives to the participant nodes while achieving security and privacy.

1.1.3 Information spread in VANETs

Vehicular ad hoc networks permit a vehicle to automatically warn nearby vehicles about its movements (braking, lane change, etc.) to avert dangerous situations. These *alert messages* only require a limited dissemination (less

than a hundred meters) but have very strong real-time requirements (they must be processed very quickly).

VANETs also allow a car to send announcements about road conditions (traffic jams, accidents, icy spots) to other vehicles so that the latter can take advantage of that information to select routes avoiding troublesome points. Such *announcement messages* require a longer dissemination range. However, their requirement of real-time processing is much less strict than in the case of alerts, so that advanced cryptography can be used to make such messages secure and trustworthy.

Privacy (anonymity and unlinkability) is a key aspect in vehicular ad hoc networks. The fact that a vehicle is equipped with communication capabilities should not render profiling its driver's habits (locations visited, driving pattern, etc.) any easier. Indeed, as noted in [Dötz06] a lot can be inferred on the driver's personality if the whereabouts and the driving pattern of a car can be tracked.

Security in car-generated announcements sent over a VANET is fundamental. It is particularly important that the system does not permit an intruder (external attacker) or a dishonest driver (internal attacker) to attack integrity by either inserting fake announcements or modifying announcements sent by others. Tampered announcements could seriously disrupt traffic or cause dangerous situations for other vehicles.

Our interest in this field focuses on the design of protocols which enable vehicles to generate and spread announcement messages compromising neither the privacy of the users nor the security of the network.

1.1.4 Resource access in social networks

In some social networks, users can specify how much they trust other users by assigning them a trust level [Ashr06, Saba06]. It is also possible to establish several types of relationships among users (for example, "colleague of", "friend of ", etc.). The trust level and the type of relationship are used to decide whether access is granted to resources and services being offered.

As pointed out in [Carm07, Mikr07], the availability of information on relationships (trust level, relationship type) has increased with the advent of the Semantic Web and raises privacy concerns: knowing who is trusted by a user and to what extent discloses a lot about that user's thoughts and feelings. See [Barn06] for an analysis of related abuses.

These privacy issues have motivated some social networks [Face08, Vide08] to enforce simple protection mechanisms, according to which users can decide whether their resources and relationships should be public or restricted to themselves and those users with whom they have a direct relationship. Unfortunately, such straightforward mechanisms result in too restrictive policies.

Regarding this topic, we focus on enabling private relationships in social networks while preserving the network functionality.

1.2 Structure of this thesis

This thesis is organized as follows.

Chapter 2 presents a state of the art of the different topics covered in this thesis. It is divided in four main sections. The first one deals with the application of security in many-to-one communications. The second one focuses on secure and private information sharing in mobile ad hoc networks. The third section reviews current work related to private and secure information

spread in VANETs. Finally, the last part of this chapter deals with privacy issues when accessing to resources in social networks.

Chapter 3 presents our contributions to secure many-to-one symbol transmission for resource-constrained devices. First, a protocol that minimizes bandwidth usage is presented. Therefore, this proposal is useful in environments where the bandwidth is a scarce resource and it is critical to make the most of it. Next, a protocol that offers secure many-to-one symbol transmission for sensor networks is presented. It provides an optimal message length and the computational cost at nodes is reduced enough to work properly on lightweight nodes. We refer to the computational capabilities on real sensor devices to prove the deployability of this proposal in real environments.

Chapter 4 presents our contributions to information sharing in mobile ad hoc networks. The first section presents an information system where the information servers are static nodes and the users who request information are mobile nodes. The main objective of such system is to give information *just in time* and *just in place* to users in an certain urban area. A typical application of this construction would be provision of touristic information. The second section presents a new scheme designed to disseminate advertisements through mobile ad hoc networks. This scheme exploits the capabilities of mobile ad hoc networks to increase the visibility of the products being offered by merchants. It outperforms current proposals in the literature. A new approach to reward nodes that collaborate in the dissemination is provided too.

In Chapter 5, our contributions to private and trustworthy information spread in vehicular ad hoc networks are presented. More precisely, we present a new system that provides secure vehicle-generated announcements

on VANETs. This scheme relies on *a priori* measures against internal attackers (vehicles in the VANET sending fake messages). It outperforms current proposals in the literature. Regarding privacy, three different variants of the system are proposed to achieve privacy without losing trustworthiness. The feasibility of this scheme is studied using simulations.

Chapter 6 presents our contributions to private resource access in social networks. More specifically, we present a new protocol which offers private relationships allowing resource access through indirect relationships without requiring a mediating trusted third party (although an optimistic trusted third party is used which only acts in case of conflict). This scheme addresses the functionality and privacy drawbacks found in current proposals in the literature. Empirical evidence is provided about the proposed protocol being scalable and deployable in practical social networks.

The concluding remarks and a summary of the results presented in this thesis can be found in Chapter 7. Some guidelines for future research are given in that chapter as well.

Chapter 2

State of the Art

In this chapter, we present the state of art of the different topics covered in our research. This chapter will serve as a basis to identify unsolved subjects that will be later addressed in this thesis.

2.1 Many-to-one information transmission

As stated in [Wolf03], the solution to implosion in many-to-one scenarios is obtained by intermediate routers combining received messages into a single message that is routed towards the base station (the root of the tree). This process is called *aggregation*. The authors in [Wolf03] present a general framework for scalable many-to-one communication where intermediate nodes collect messages from their children, aggregate them and send a single aggregated message up to their parent. In this way, the base station receives a single message containing all the readings from the leaves. This solution is scalable (permitting an unlimited amount of senders) as long as aggregated data do not grow in size. Two scenarios are then possible:

- *Lossy aggregation.* In this case, the message output by aggregation

contains less information than the set of messages input to aggregation. Thus, the size of the output can stay the same as the size of each input. Some examples of lossy data aggregation are:

- If data is a temperature, different temperatures can be aggregated by computing their average. Information loss comes from the fact that the base station will not know the temperature obtained by each node but only the average of all readings.
- If data is a counter, different counters can be aggregated by addition. Information loss comes from the base station not being able to find out the exact contribution by each node.
- If data sent is a binary value indicating an alarm, it can be aggregated using a logical OR operation. The base station will know an alarm has been raised but not its exact origin.

On the whole, lossy approaches can not be used in scenarios where the root must know the specific data sent by each leaf.

- *Lossless aggregation.* This situation occurs when no information loss is affordable during aggregation. It happens in applications where the root multicasts a data request to the leaves and the leaves react by sending one q-ary symbol each (data sent by each leaf can be modeled as an integer ranging from 1 to q). At the end of the process, the root knows which symbol was transmitted by each leaf. In this case, the only possibility left is for leaves to use a message length such that all information they transmit can be aggregated in a single message of that length (the message reaching the root). This implies that the actual informational content transmitted by leaves will be less than the bitlength of the messages they use.

The framework presented in [Wolf03] works fine in these two scenarios. However, it does not address security. This fact represents a major drawback which disqualifies it when security requirements arise.

Some proposals for secure many-to-one communications in both scenarios exist in the literature. We summarize them in the following two subsections: *Secure many-to-one lossy transmission* and *Secure many-to-one lossless transmission.*

2.1.1 Secure many-to-one lossy transmission

Few researchers have proposed solutions which provide security in this kind of lossy communications. In [Przy03] the authors present a framework for designing secure data aggregation protocols. They propose concrete protocols within this framework for securely computing the median, securely finding the minimum and maximum values, securely estimating the network size and securely computing the average of measurements. This framework assumes the existence of special nodes called aggregators which receive the readings from the sensor nodes and aggregate them. The authors state that a user can verify that a certain aggregation given by an aggregator is a good approximation of the true value even when such aggregator and a fraction of the sensor nodes are corrupted. However, there are some shortcomings in this framework:

- It only offers data confidentiality against external attackers eavesdropping the path from a particular sensor node to the aggregator. Since aggregators must know the sensor readings that they are aggregating and their sources, when an aggregator becomes compromised, the confidentiality of all the messages which traverse such node becomes compromised too.

- The framework presented in [Przy03] is designed to work with only one aggregator. The authors admit that in case of too large a sensor network, an aggregator alone may not be capable of handling the whole network. In this case, they propose to use several aggregators to perform a process named *hierarchical aggregation*. Even though this solution works with some functions like Min/Max and average computation, there are other functions which can not be treated in this way. The authors leave the research on new aggregation types as future work.

- This proposal assumes that a static network is used. Sensor nodes and aggregators share preloaded secret keys and they must be deployed in a deterministic way. Therefore, this framework is not feasible in scenarios where the nodes are randomly deployed.

In [Jadi04] the authors present a secure aggregation protocol that provides confidentiality as well as integrity guarantees. This protocol aggregates encrypted data directly, without requiring decryption at intermediate nodes. This preserves the confidentiality of the data while they traverse the network towards the base station. However, this protocol only enables the intermediate nodes to compute an addition of the received sensor readings. Therefore, such a protocol is unsuitable for certain kinds of queries like Min/Max. The authors leave this issue as part of their future work. Besides, the authors admit that a collusion by a certain parent and one of its child nodes can misrepresent the readings of the whole subtree without being detected. In addition to that, the presented protocol does not work in scenarios where the nodes are randomly deployed.

[Dimi05] and [Dimi06] deal with how to set up the network when the sensor nodes are deployed randomly. Even though they incentivize the use of aggregation in intermediate nodes, such proposals do not present any concrete

protocols to aggregate the sensor readings and the author leaves this as an open issue.

2.1.2 Secure many-to-one lossless transmission

Regarding security in many-to-one lossless communications. There are some schemes in the literature which address this scenario. Such proposals can be divided into two categories described below: *secure acknowledgment* and *secure symbol transmission*.

Secure acknowledgement

These schemes provide the root with an undeniable and unforgeable proof that a certain set of leaves have received a specific content. The information sent by the leaves to the root is unary in the sense that, after receiving a piece of data, every leaf will either respond with a positive acknowledgement (a digital signature) in case of correct reception or will stay silent otherwise.

The systems proposed in [Nico04] and [Cast05] fall into this category. The former uses the multisignature scheme in [Bold03] constructed over a Gap Diffie-Hellman group (GDH) [Bone01]. The latter is a construction whose security rests on the hardness of the discrete logarithm problem. Both solutions provide non-repudiation and are scalable ($O(n)$ message length) as long as the set of acknowledging leaves remains stable.

These systems only provide non-repudiation; other security properties are not addressed. For instance, the root is unable to distinguish a voluntary non-transmission from malicious erasure of acknowledgements by intruders. The authors in [Nico04] and [Cast05] leave this issue for future work. Thus, integrity is not ensured. Confidentiality is not achieved either since any

intruder listening to the communication can ascertain which leaves are acknowledging and which are not.

Secure symbol transmission

Such schemes assume a tree communication model where the root multicasts a data request to the leaves. Upon reception of this request, the leaves react by sending one q-ary symbol each. These messages will be aggregated by intermediate nodes. From the received message, the root will obtain the symbol sent by each leaf.

It can be proven that symbols sent by n leaves can not be aggregated in a message whose length is below $O(n)$ when all symbols have the same probability of being sent. Current research in secure symbol transmission focuses on designing systems whose actual length of messages is as short as possible (within the $O(n)$ length class). Note that this fact does not permit an unlimited amount of senders.

In [Domi04] the authors propose a system using super-increasing sequences and additive privacy homomorphisms. The length of messages is $O(n)$, where n is the number of leaves of the multicast tree. If implemented using the Okamoto-Uchiyama cryptosystem [Okam98] for binary transmissions the message length asymptotically tends to $6n$. The scheme is easily extensible to accommodate q-ary alphabets with message length tending to $3tn$ (where $q \leq 2^t - 1$).

The proposal in [Sebe07a] reduces the message length with respect to the scheme presented in [Domi04] for biased binary communication –*i.e.*, where the probability of leaves transmitting a '1' symbol is less than the probability of their transmitting a '0' symbol. This scheme offers an $O(k \log k \log n)$ message length with n being the number of leaves and k being an upper

bound on the number of leaves that wish to transmit simultaneously the least likely symbol. Both systems provide confidentiality, authentication and integrity. Non-repudiation is not provided.

In spite of their bandwidth efficiency, both proposals present a high computational cost. Both use additive public-key privacy homomorphisms, whose cleartext message length grows like $O(n)$ for [Domi04] and grows like $O(k \log k \log n)$ for [Sebe07a]. The costly cryptographic operations on long messages required by these schemes render them ill-suited for implementation on resource-limited hardware like the sensor nodes used in sensor networks.

Regarding integrity, both systems permit data corruption to be detected but identifying the corrupting nodes is not straightforward. This must be done using a tracing procedure described in [Sebe07a] (which can also be applied for [Domi04]) in which the root traces and identifies corrupting nodes.

In Section 3.1 a scalable tree-based protocol for secure many-to-one symbol transmission is proposed. This protocol saves more bandwidth and it is computationally simpler than previous proposals in the literature. However, this new scheme is still not enough lightweight to work properly in sensor nodes. Section 3.2 presents a new protocol for secure many-to-one symbol transmission in which nodes are only required to perform very simple operations. This makes it suitable for implementation in resource-constrained scenarios such as sensor networks. Both new schemes provide their own methods designed to identify the corrupting nodes when data corruption is detected. Such procedures are more efficient than the one presented in [Sebe07a] for this purpose.

2.2 Information sharing in MANETs

As stated previously, successful information sharing models in MANETs must address issues related to data eavesdropping, data tampering, packet dropping and selfish behaviour. The first three attacks can be addressed by applying cryptography-based secure communications. The last one is addressed by giving incentives to the collaborative nodes of the network. In addition to that, communication protocols for MANETs must be privacy-preserving. Provision of incentives while offering security and privacy to the users is a challenging task.

We next introduce the existing proposals in the literature regarding privacy, security and incentives in MANETs.

2.2.1 Privacy, security and incentives

Preservation of user privacy sometimes contradicts security requirements. For example, a system offering services needs users to authenticate themselves to be sure it will receive a correct payback. Another example occurs when a certain user has an inappropriate behavior in the network. The system has to identify the intruder in order to take proper measures. Measures to secure these situations may affect the user's privacy.

In [News04] a system is proposed where all nodes are registered. As a result, the system is secure against external or internal attackers but this approach does not respect the privacy of the users.

User authentication and privacy in MANETs are addressed in [Weim04]. In this work the authors present a protocol that allows nodes in a MANET to recognize each other when meeting again. This scheme provides provably

secure authentication against passive adversaries and secure message authentication against active adversaries. Besides, it provides privacy while keeping immutable and non-migratable identities. The shortcoming of this proposal is that users can freely change their identities, which can be exploited by dishonest users to disrupt the system. Therefore, this protocol is not suitable for real environments with active adversaries.

Privacy issues become worse when incentives should be given to the collaborative nodes of the network. As stated in [Vass03], the motivation of users to participate is a crucial factor for the success of a system designed for wireless ad hoc networks. However, providing motivation to nodes who offer services to other users implies the need for a secure and private way for collecting the rewards from the served users.

Secure electronic payment is a profusely studied research topic. From electronic money to e-coupons [Blun05], there are several electronic payment methods suitable for mobile devices. Nevertheless, for the specific case of secure and private incentive-based schemes, the literature is rather scarce.

[Raja05] and [Vish03] propose incentive-based schemes where the network nodes have an account and the content provider gives them credit depending on the information they have uploaded. The network nodes can use their credit to increase their download rate or change it for money. Nonetheless, these proposals are not designed for a mobile ad hoc network, and the security is only focused toward the protection of the copyrighted content. Thus, the credit of the network node can be tampered with. In addition to that, privacy issues are not considered.

In [Pan07], the authors propose a lightweight and cheat-resistant micropayment scheme to stimulate and compensate collaborative peers that sacrifice their resources to relay packets for other peers. This scheme focuses on

providing a secure and stable channel to exchange data between two peers within an ad hoc network. Intermediate nodes are incentivized to keep this channel operative. Even though this proposal provides security to information provision services in MANETs, it does not preserve the privacy of the users. A certain node involved in this scheme relies on its identity to receive information and communicate with other peers. In addition to that, its identity is related to its reputation (*e.g.* cooperativeness in relaying) and wealth (*e.g.* collected credits for its cooperation) in the system. If a peer is found to be malicious, either persistently or opportunistically, such node can be excluded from the system by identity blacklisting.

A different method for sharing information in MANETs is introduced in [Stra04]. In this work, the authors present AdPASS. This is a new system to disseminate information in MANETs exploiting the mobility property inherent in this kind of networks. The authors focus their work in an M-commerce[1] application which spreads digital advertisements among interested users present in a MANET. Each user specifies her interests in a profile that is stored in her mobile device. When a certain user gets an advertisement of another user's interest, she spreads it every time she finds new interested users around. If a customer uses the acquired advertisement to buy something in the source shop, all the users who have cooperated in the dissemination of the advertisement will receive some bonus. Such bonus can later be exchanged by goods in the source shop. Even though this scheme is supposed to provide security and privacy to the users who disseminate advertisements, it is weak against dishonest nodes which cooperate to steal bonuses from other users. Another shortcoming of this proposal is that the

[1] *M-commerce stands for electronic trading of goods and services made through mobile devices.*

authors only explain how to get the bonus points but they do not mention how such points are later spent. This issue must be addressed since privacy could be compromised at this point. Besides, this approach requires the users to register themselves to a trusted authority named *mediator* which acts as *anonymizer* and keeps track of the user's accumulated bonus points. We claim that a system which is designed to work in MANETs should not require the presence of a trusted third party (TTP).

In addition to that, the bonus points scheme in AdPASS offers no guarantees of fairness: even though a reasonable behavior can be expected, the fact is that each user disseminating an advertisement can take as many points as she wishes, regardless of how many she actually deserves. Worse yet, collusions are conceivable where colluders exclude other users from dissemination in order to monopolize bonus points. AdPASS must definitely be repaired to thwart those roguish attitudes. Last but not least, the total number of bonus points assigned by the merchant to an advertisement is a *de facto* upper bound on the number of feasible transfers to new disseminators: due to the limited range of MANET nodes, this implies some limitation in the geographical dissemination range and the sales potential.

On the whole, several proposals in the literature use incentives to avoid node misbehavior. However, applying privacy and security to incentivized information sharing in mobile ad hoc networks has not been investigated enough.

In Section 4.1 an architecture for a peer-to-peer mobile ad hoc network offering distributed information provision is presented. The proposed architecture is specified as a protocol suite taking security, privacy and incentive aspects into account. Section 4.2 describes a secure and private scheme to

disseminate advertisements in mobile ad hoc networks where collaborating nodes are incentivized by giving them e-coins.

2.3 Information spread in VANETs

As explained later, VANETs demand protocols which enable vehicles to generate and spread announcement messages without compromising the privacy of the users nor the security of the network.

Security against insertion of fake announcements by external attackers is easy to achieve using well-known cryptographic authentication techniques (digital signatures or message authentication codes). Such techniques require the sender of a message to access some secret key material only available to legitimate, registered users —and therefore unavailable to external attackers.

Dealing with internal attackers is a thornier issue. The reason is that legitimate system users, and thus internal attackers, have access to the secret key material required to send authenticated fake messages (for instance, to announce a false traffic jam with the aim of diverting traffic from a certain area where some kind of crime is being committed). Countermeasures against fake messages from internal attackers fall into two classes: *a posteriori* and *a priori*.

2.3.1 A posteriori countermeasures against fake messages

A posteriori countermeasures consist of taking punitive actions against users who have been proven to have originated fake messages (*e.g.* the offenders can be banished from the network). These countermeasures in anonymous systems require the presence of a trusted third party able to revoke the key

material of such dishonest users. In this way, they will be excluded from the system.

Digital signatures have been extensively used in most of the protocols that offer *a posteriori* countermeasures: from plain digital signatures [Raya06a, Raya07a, Raya07b, Armk07] until more sophisticated distributed signatures, such as group signatures in [Guo07] or ring signatures in [Lin07]. The latter paper and [Gama06] also consider ID-based ring signatures.

2.3.2 A priori countermeasures against fake messages

A priori countermeasures attempt to prevent the generation of fake messages. In this approach, a message is not considered valid unless it has been endorsed by a number of vehicles above a certain threshold. Those vehicles must be in a position to confirm what is reported in the message: for a traffic jam announcement, other jammed vehicles are potential endorsers (automatically or after intervention of their drivers); for an "icy road" message, nearby vehicles whose traction system has detected slippery ground can be automatic endorsers. This approach is based on the assumption that most users are honest so that they will not endorse any message containing false data.

Under this approach, the risk that a collusion of dishonest vehicles reaches the size necessary to generate fake messages always exists. The natural strategy against collusions is to choose a threshold sufficiently high so as to render successful collusions unlikely. However, this threshold should not be so high that it prevents honest vehicles from sending true announcements in situations with a low density of vehicles.

The *a priori* approach is compatible with driver privacy: since false announcements are thwarted without resorting to punitive actions, unconditional vehicle anonymity is allowable (in contrast, *a posteriori* countermeasures assume that offenders are identified and punished).

The use of a honest majority to prevent generation of fake messages has previously been proposed in [Goll04, Parn05, Oste07, Raya06b]. A brief discussion of those papers is next given.

In [Goll04] a framework is presented to validate received data in VANETs. In this approach, a vehicle receives alerts from different neighbors and compares them in order to infer the correctness of a certain event. This scheme suffers from high communication overhead due to the lack of aggregation techniques. Besides, the proposed framework has not been empirically tested. The paper [Parn05] presents a contribution that remains quite vague: VANET security issues are identified, some security primitives are enumerated, but no complete protocol is actually described. In [Oste07], a system that evaluates the plausibility of received danger warnings is proposed. This system estimates the trustworthiness of a reported hazard by taking a vote on the received danger messages. The paper provides a simulative analysis of different voting schemes, but privacy remains unaddressed and security is not completely covered. Finally, [Raya06b] describes a detailed protocol deployable in real VANET environments (the authors show this via simulation) which systematically deals with security threats and reduces communication overhead by aggregating messages.

According to the above discussion, [Raya06b] seems the most competitive scheme in the literature on the *a priori* approach, so we concentrate on it in what follows. That paper presents three variants offering *a priori* countermeasures against fake messages: *concatenated signatures*, *onion signatures*

and *hybrid signatures*.

In the variant based on *concatenated signatures*, a vehicle generates an announcement and sends it, its signature and its public-key certificate to a nearby car which will endorse it by computing its own signature on it. This new signature and the corresponding public-key certificate will be appended to the frame that will be retransmitted to the next vehicle. An announcement is considered valid after it has been endorsed by at least the number of vehicles determined by the threshold. This approach has several drawbacks:

- It does not offer unlinkability since different signatures made by the same user can be linked through the public key that verifies them. Anonymity is however feasible by using pseudonyms.

- Announcement generation is delayed due to the sequential communication pattern (the delay is proportional to the number of endorsing vehicles).

- It requires the verifier to check several signatures upon receiving an announcement (as many verifications as vehicles have endorsed the message). These verifications involve checking the validity of public-key certificates and probably revocation lists as well.

Therefore, there is room for improvement both in terms of privacy and efficiency (communication and computation costs).

The variants based on onion signatures and hybrid signatures are similar and designed to reduce the overall message length. Both variants use the so-called *oversignatures*: instead of appending its signature, each new endorsing car signs the signature by the previous endorsing car (this is called oversigning). In an oversignature, a verifier can check the last endorser's signature, but not the signatures by the previous endorsers. Since this is a

serious design flaw, we will only consider the *concatenated signatures* variant for comparison in the rest of this paper.

In Section 5.2 a new scheme is presented following the *a priori* protection paradigm that reduces the verification cost of endorsed messages to one signature verification. In this proposal, vehicles volunteering to generate and/or endorse trustworthy announcements do not have to sacrifice their privacy (anonymity and unlinkability).

2.4 Resource access in social networks

In the introduction of this thesis, we have pointed the existence of some social networks [Face08, Vide08] which use too restrictive protection mechanisms to preserve their users' privacy when performing a resource access. Nonetheless, those are not the only proposals in the literature which address this topic.

In [Carm06], a more flexible access control scheme is described, whereby a *requestor* can be authorized to access a resource even if he has no direct relationship with the *resource owner*, but he is within a specified depth in the relationship graph. *Access rules* are used, which specify the set of *access conditions* under which a certain resource can be accessed. Access conditions are a function of the relationship type, depth and trust level. Relationship certificates based on symmetric-key cryptography are used by a requestor to prove that he satisfies some specific access conditions. To access resources held by a node with whom the requestor has no direct relationship, the requestor retrieves from a central node the chain of relationship certificates along the path from the resource owner to himself. Clearly, the central node is a trusted third party, as it knows the relationships of all nodes in the

network.

In [Wang06] a mechanism to protect personal information in social networks is described where nodes in the network are anonymous and cannot be linked to specific users; in contrast, the data and the relationships are public, which might facilitate user re-identification.

An innovative privacy-preserving approach is described in [Carm07] which leans on the access model in [Carm06] and focuses on relationship protection: a user can keep private that he has a relationship of a given type and trust level with another user. Relationship certificates are encrypted and are treated like a resource in their own right: access to a certificate is granted using a *distribution rule* for that certificate, where the *distribution conditions* to be satisfied by users wishing to access the certificate are specified. If a user satisfies the distribution rule for a certificate, he receives the corresponding symmetric *certificate key* allowing him to decrypt the certificate. In [Carm07] a rather complex scheme is proposed to manage and distribute certificate keys. Encrypted certificates are stored at a central node; due to encryption, the central node does not have access to the cleartext certificates, so it does not need to be trusted in this respect. However, the central node needs to be trusted in the following aspects:

- *Trust level computation* when several relationship certificates are chained (indirect relationship between a resource requestor and a resource owner).

- *Certificate revocation enforcement* when a relationship ceases to exist (the central node must maintain a certificate revocation list and inform the other nodes about new revocations).

In [Domi07] a protocol is proposed which overcomes the shortcomings

detected in [Carm07]. Specifically, the author presents a public-key protocol which achieves relationship protection without the presence of a central node working as trusted third party. In addition to that, this protocol avoids revealing the content of relationships to the resource requestor and substantially simplifies relationship revocation. Nevertheless, this scheme has some shortcomings that we next summarize:

- For each resource access, a user tries to get the backing of the nodes with whom he is related. If a related node is temporarily unreachable or refuses to collaborate, it is hoped that other nodes related to the requestor will be available to act as intermediate nodes. However, a user with a small number of relations is likely to stay isolated at certain periods of time (*e.g.* early in the morning). This issue is an open problem which must be addressed.

- The protocol in [Domi07] prevents the resource requestor from seeing any of the relationship certificates that will be used by the resource owner to decide whether the requestor is granted access. However, the resource owner learns the relationships, and their trust level, between the users who collaborate in the resource access. This represents a major privacy toll which would justify that some intermediate nodes might refuse collaboration. This fact also has implications for the previous point explained above: nodes which refuse to collaborate add to nodes which are unreachable and both categories disrupt in the same way the normal network operation.

In Section 6.2 a new protocol for resource access in social networks is presented. This proposal offers the same features of [Carm07] and [Domi07]

while providing a solution which addresses the drawbacks left open in [Domi07].

Chapter 3

Efficient and secure many-to-one symbol transmission

In Section 2.1 we pointed out the need to design efficient protocols providing secure many-to-one symbol transmission for sensor networks. In this chapter, we present our contributions to this field.

Sensor networks are formed by devices with limited computational capabilities and limited battery power. Therefore, this kind of networks need protocols which provide reduced bandwidth usage and low computational cost at the sensor nodes.

Section 3.1 presents a scheme for many-to-one symbol transmission that has been published in [Sebe07b]. It provides an optimal message length. In this way, bandwidth usage is reduced to the minimum. According to that, this proposal is useful in environments where the bandwidth is a scarce resource and it is critical to make the most of it. It also provides immediate detection of corrupted messages. This scheme uses multisignatures over Gap

Diffie-Hellman (GDH) groups [Bold03]. Note that these cryptographic operations may not be suitable for implementation in resource-limited networks like sensor networks.

Section 3.2 presents the first proposal in the literature that offers secure many-to-one symbol transmission for sensor networks. It has been published in [Viej08]. This scheme is based on [Sebe07b]. It also provides an optimal message length but replaces the use of GDH cryptography with hash functions. In this way, computational cost at nodes is reduced. As a result, this proposal is suitable for resource-constrained devices, which are quite common in sensor networks. This scheme does not permit immediate detection of corrupted messages. This detection is performed using an *a posteriori* tracing algorithm.

3.1 Secure many-to-one communications based on GDH multisignatures

In this section, we introduce a scalable and secure protocol for many-to-one symbol transmission that offers an optimal message length and is computationally simpler than previous proposals in the literature [Domi04, Sebe07a]. In addition to that, nodes can immediately check the correctness of received messages and detect data corruption without requiring any extra error tracing procedure. In previous proposals, message corruption was detected by the root of the tree communication model. Identification of dishonest nodes was done using a tracing algorithm. Last but not least, computational cost at nodes is lower than in previous proposals. This is proven in Section 3.1.5.

As stated previously, the construction we present uses multisignatures over a Gap Diffie-Hellman group [Bold03]. Next, we briefly introduce its

mathematical background. Later, we describe our protocol in detail.

3.1.1 Multisignatures over Gap Diffie-Hellman groups

A Gap Diffie-Hellman (GDH) group G is an algebraic group of prime order q for which no efficient algorithm can compute g^{ab} for random $g^a, g^b \in G$, but such that there exists an efficient algorithm $D(g^a, g^b, h)$ to decide whether $h = g^{ab}$. Let 1_G be the neutral element of G. GDH groups are suitable for public-key cryptography. The secret key is a random value $x \in \mathbb{Z}_q$ and its corresponding public key is $y \leftarrow g^x$. The signature on a message m is computed as $\sigma \leftarrow \mathcal{H}(m)^x$ ($\mathcal{H}$ is a cryptographic one-way hash function). The validity of a signature can be tested by checking $D(y, \mathcal{H}(m), \sigma)$.

GDH groups are convenient to compute multisignatures. Given two signatures of the same message m under two different public keys y_1, y_2, a signature of m under the combined public key $y \leftarrow y_1 y_2 = g^{(x_1+x_2)}$ can be obtained as $\mathcal{H}(m)^{x_1}\mathcal{H}(m)^{x_2} = \mathcal{H}(m)^{x_1+x_2}$.

3.1.2 General assumptions

Our protocol assumes a tree communication model in which the the root is the final receiver, internal tree nodes are reverse multicast routers and the leaves correspond to senders. The root S has a private key x_S and its corresponding public key $y_S \leftarrow g^{x_S}$. This public key is accepted by all the nodes in the tree. Each leaf U_i has several private/public key pairs. The public keys are accepted as valid by intermediate routers and the root. Each node in the multicast tree knows its parent node and the public keys of nodes belonging to the subtree rooted at it. Each node also knows the public key of the root.

3.1.3 Reverse bit transmission

In this section we detail our proposal for binary communication, where each leaf U_i transmits a "0" or "1" bit (denoted by b_i). We assume the multicast tree contains n leaves U_i, $1 \leq i \leq n$. Each leaf U_i has two secret keys $x_{i,a}$ and $x_{i,b}$. Its corresponding public keys are $y_{i,a} \leftarrow g^{x_{i,a}}$ and $y_{i,b} \leftarrow g^{x_{i,b}}$. We also require each leaf U_i to share a secret key K_i with the root. This value can be agreed upon by using the Diffie-Hellman key exchange protocol. In this way, leaf U_i obtains K_i from one of its private keys and the root's public key, that is, $K_i = (y_S)^{x_{i,a}}$; the root can also obtain K_i by computing $K_i = (y_{i,a})^{x_S}$.

The protocol works as follows:

1. Challenge. The root multicasts to the leaves a challenge consisting of a random value v (v may include a description on the requested information).

2. Message generation.

   (a) Upon receiving v, each leaf U_i computes a pseudo-random bit $c_i \leftarrow lsb_1(\mathcal{H}(v||K_i))$, where $lsb_1(\cdot)$ is a function returning the least significant bit of its argument.

   (b) If $c_i \oplus b_i = 1$ then U_i computes $\sigma_i := \mathcal{H}(v)^{x_{i,a}}$.
   If $c_i \oplus b_i = 0$ then U_i computes $\sigma_i := \mathcal{H}(v)^{x_{i,b}}$.

   (c) If $c_i \oplus b_i = 1$ then U_i generates a $2n$-bit sequence I_i so that its $2i$-th bit is "1". The rest of bits are set to "0".
   If $c_i \oplus b_i = 0$ then U_i generates a $2n$-bit sequence I_i so that its $(2i-1)$-th bit is "1". The rest of bits are set to "0".

   (d) U_i sends the pair (I_i, σ_i) up to its parent node.

3. MESSAGE AGGREGATION. An intermediate router R or the root S receives messages from its child routers/leaves and does the following:

   (a) For each received pair (I_j, σ_j):

      i. Let $i := 1$. Let $y := 1_G$.

      ii. While $i \leq n$ loop

         - If $I_j[2i] = 1$ and $I_j[2i-1] = 0$ then $y := y \cdot y_{i,a}$
         - If $I_j[2i] = 0$ and $I_j[2i-1] = 1$ then $y := y \cdot y_{i,b}$
         - If $I_j[2i] = 1$ and $I_j[2i-1] = 1$ then $ERROR$ [1]
         - $i := i + 1$

      iii. It checks $D(y, \mathcal{H}(v), \sigma_j)$. If this check fails, then $ERROR$.

   (b) Once all expected messages $\{(I_j, \sigma_j)\}_j$ have been received and checked (for the sake of simplicity, we describe the protocol assuming no errors were found), R or S aggregate them by computing $I = \bigvee_j I_j$ ($\vee$ denotes the bit-wise OR operation) and $\sigma = \prod_j \sigma_j$.

   (c) If the aggregating node is an intermediate node R, it sends (I, σ) up to its parent node. Else, if it is the root S this is the final aggregated message.

4. SYMBOL EXTRACTION. From the final aggregated message (I, σ), the root S obtains the bit sent by each leaf as follows:

   (a) Let $i := 1$

   (b) While $i \leq n$ loop

      - Compute $c_i \leftarrow lsb_1(\mathcal{H}(v||K_i))$
      - If $I[2i] = 1$ and $I[2i-1] = 0$ then $d_i := 1$ and $b_i := d_i \oplus c_i$

[1] Section 3.1.6 describes how to handle erroneous situations.

- If $I[2i] = 0$ and $I[2i-1] = 1$ then $d_i := 0$ and $b_i := d_i \oplus c_i$
- If $I[2i] = 0$ and $I[2i-1] = 0$ then $b_i := NULL$
- $i := i + 1$

(c) Return $B = (b_1, \ldots, b_n)$

Note. No verification of the signature σ is needed during the extraction step, because σ is the aggregation of signatures σ_j which have been verified at each aggregation step (in the last aggregation step, verification has been carried out by the root itself).

3.1.4 Security analysis

We next analyze how our proposal provides confidentiality, authentication, integrity and non-repudiation.

Confidentiality

The confidentiality property refers to the fact that only the root should be able to obtain vector $B = (b_1, \ldots, b_n)$ containing the bit transmitted by each leaf. An intruder eavesdropping messages of the form (I, σ) can determine for each leaf U_i located below the sniffing point in the tree whether the leaf transmitted $\sigma_i = \mathcal{H}(v)^{x_{i,a}}$, $\sigma_i = \mathcal{H}(v)^{x_{i,b}}$ or did not transmit by observing I (exactly the bits at $I[2i]$ and $I[2i-1]$).

From knowledge of σ_i the intruder is able to determine $c_i \oplus b_i$. But since c_i is only known to U_i and the root (it is computed from the challenge v and the shared secret key K_i), nobody but them is able to determine the transmitted value b_i.

Note that an intruder can determine which leaves did not transmit. In applications where this fact causes information leakage, non-transmission

should not be permitted.

Authentication

This property requires that intruders cannot generate false messages that will be accepted as valid by the system. The creation of a message that will be accepted as authentic coming from U_i requires knowledge of its private key $x_{i,a}$ or $x_{i,b}$. This is because the message sent by U_i includes a signature σ_i over $\mathcal{H}(v)$ computed from $x_{i,a}$ or $x_{i,b}$. As long as secret keys are not compromised and the signature scheme is unforgeable (a valid signature can only be computed if the secret is known) the system provides authentication. The use of a different challenge v at each execution prevents replay attacks.

Integrity

This property requires being able to detect substitution or suppression of messages by an intruder. Given a message (I, σ), the field σ is a multisignature on $\mathcal{H}(v)$. Without loss of generality, let us take the case of a leaf U_i whose message has been aggregated into (I, σ) and assume that U_i transmitted $\sigma_i = \mathcal{H}(v)^{x_{i,a}}$. Further, assume that the value $\sigma_i = \mathcal{H}(v)^{x_{i,a}}$ is known to an attacker who could have obtained it by capturing the first message sent by U_i.

An attacker wishing to replace U_i's contribution with $\sigma_i' = \mathcal{H}(v)^{x_{i,b}}$ needs to replace σ with σ' so that $\sigma' = \sigma\mathcal{H}(v)^{x_{i,b}} \left(\mathcal{H}(v)^{x_{i,a}}\right)^{-1}$. If this was possible, an attacker able to compute σ' would get $\mathcal{H}(v)^{x_{i,b}} := \frac{\sigma'\sigma_i}{\sigma}$ which is a signature on $\mathcal{H}(v)$ which would be validated using the public key $y_{i,b}$. This would contradict the unforgeability property of the GDH signature. In this sense, the system provides integrity against malicious alteration of the value b_i sent by a given leaf.

An intermediate node could dishonestly decide to suppress and not aggregate a message received from some of its child nodes. This fact would be interpreted by the root as a non-transmission. Also, the contribution by U_i could be suppressed by an attacker who knew the value σ_i ($\mathcal{H}(v)^{x_{i,a}}$ or $\mathcal{H}(v)^{x_{i,b}}$). This can be done by computing $\sigma' = \sigma(\sigma_i)^{-1}$ (the corresponding alteration of I is trivial).

In order to avoid these suppression attacks, the protocol should not permit non-transmissions. In this way, if the root gets nothing from a leaf, a suppression attack is signalled.

Non-repudiation

This property requires that no leaf be able to deny having sent a given value that has been received by the root. A valid message (I, σ) reaching the root contains a value σ that is a multisignature on $\mathcal{H}(v)$. The non-repudiation property of the multisignature scheme guarantees that each leaf having contributed to the signature cannot deny having signed under the private key corresponding to one of its two public keys $y_{i,a}$ or $y_{i,b}$. Therefore, a leaf U_i cannot repudiate the value $c_i \oplus b_i$ she sent. Deriving non-repudiation on b_i from $c_i \oplus b_i$ requires prior declaration by U_i of the procedure used to obtain c_i. This procedure must be later reconstructable in front of a third party.

3.1.5 Performance analysis

Performance will be analyzed in terms of message length and computational cost. We compare our protocol with [Domi04]. On the other hand, [Sebe07a] is not included in this comparison because it was designed for biased scenarios and we are focused on non-biased scenarios.

Message length

In our proposal, the length of messages stays constant in the way from the leaves towards the root. A message consists of the pair (I, σ). The bitlength of component I is $2n$ (being n the number of leaves) while the component σ is a multisignature constructed over a Gap Diffie-Hellman group [Bone01]. This group can be constructed over non-supersingular elliptic curves to get a bitlength of approximately 170 bits which provides a security level similar to 320-bit DSA signatures or 1024-bit RSA signatures [Bone01]. Thus, messages have a bitlength $2n + O(1)$. For large values of n, this length tends asymptotically to $2n$. This improves on the message length offered by [Domi04], which tends asymptotically to $6n$.

Computational cost

The calculations performed by the protocol can be classified into four categories: message generation, message verification, message aggregation and data extraction. We next quantify the computation in each category.

Generation. Generation of (I_i, σ_i) by leaf U_i takes time $O(n)$ to generate the binary sequence I_i plus the time to compute the signature σ_i. Since this latter time does not depend on n, we take it as $O(1)$ in our analysis.

Verification. An intermediate node receives and checks messages $\{(I_j, \sigma_j)\}_j$ from its child nodes. For each (I_j, σ_j), the node computes y and then checks the validity of the multisignature σ_j. Computation of y requires one operation over the GDH group for each leaf that contributed to the message. The verification time of the signature does not depend on n, so we take it $O(1)$. Since there are $O(n)$ leaves in the tree, the overall amount of multiplications spent by one node computing the y's for all

$\{(I_j, \sigma_j)\}_j$ is at most $O(n)$. The amount of signatures to be verified depends on the number of child nodes of the intermediate router. In any case, this amount cannot grow faster than $O(n)$.

Aggregation. Aggregation of $\{(I_j, \sigma_j)\}_j$ into (I, σ) requires at most $O(n)$ bitwise OR operations over $O(n)$-long messages during the computation of I and at most $O(n)$ operations (each one with cost $O(1)$) over the GDH group to compute the new multisignature. This results in a maximal $O(n^2)$ cost.

Extraction. Finally, the extraction of vector $B = (b_1, \ldots, b_n)$ by the root node is done by processing component I in time $O(n)$.

Table 3.1 compares our system with respect to [Domi04]. The cubic cost ($O(n^3)$) of message generation and data extraction in [Domi04] is due to the encryption of $O(n)$ long messages using the Okamoto-Uchiyama homomorphic cryptosystem.

Table 3.1: Performance comparison with [Domi04]

	Our proposal	[Domi04]
Message length (for $n \uparrow\uparrow$)	$2n$	$6n$
Cost of message generation	$O(n)$	$O(n^3)$
Cost of message aggregation	$O(n^2)$	$O(n^3)$
Cost of data extraction	$O(n)$	$O(n^3)$

3.1.6 Error handling

Upon message reception, intermediate nodes perform several checks on the messages (I_j, σ_j) received from their child nodes prior to composing the aggregated message they will transmit. The checks that are always performed

are:

- Check that I_j does not contain $I_j[2i] = I_j[2i-1] = 1$ for any U_i (message generation does not permit this situation).
- Check that multisignature σ_j is consistent with the aggregated public key y computed from I_j.

If some of the above checks fail, the node will consider its sender (one of its child nodes) liable. This is because the child node either ought to have detected and reported these problems when performing its checks (if it was an intermediate node) or is causing the problems itself. In particular, if the child node is a leaf, it should have constructed an error-free message. Upon identification of a disrupting node, appropriate measures are taken against it (for instance, removal of the node from the multicast tree).

If non-transmission by leaves is not permitted, some additional requirements arise:

- First of all, an intermediate node has to receive one message (I_j, σ_j) from each of its children. If some of them are missing this will be interpretated as a malicious non-transmission by the corresponding children.
- Each intermediate node needs to know the list of leaves present in the subtree rooted at each of its child nodes. When checking each message (I_j, σ_j), it needs to check that all leaves present in this subtree are contributing. If this is not the case, this node will consider its sender child liable for having suppressed such contributions.

Note that neither the reception of a corrupted message nor a non-reception may be caused by the sender, but by an attacker disrupting the communication link between the sender and the receiver. In any case, the receiver

cannot distinguish between both situations. The receiver simply perceives that messages coming from that child are not reliable any more; upon this, the receiver can take the appropriate measures.

3.1.7 Generalization to q-ary transmission

The system can easily be generalized from binary to q-ary communications. We will represent each symbol from the q-ary alphabet by a different integer from the set $\{1, \ldots, q\}$. First of all, the smallest integer t such that $q \leq 2^t - 1$ is chosen. Each leaf U_i has t secret keys $x_{i,1}, \ldots, x_{i,t}$, with their corresponding public keys $y_{i,1} \leftarrow g^{x_{i,1}}, \ldots, y_{i,t} \leftarrow g^{x_{i,t}}$ accepted as valid by the intermediate routers and the root. Like in the binary protocol, the root S shares a secret key K_i with each leaf.

The generalized protocol is as follows:

1. Challenge. The root multicasts a challenge consisting of a random value v (v may include a description on the requested information.).

2. Message generation.

   (a) Upon receiving the challenge v, each leaf U_i computes a pseudo-random t-bit sequence $(c_1, \ldots, c_t) \leftarrow lsb_t(\mathcal{H}(v||K_i))$, where $lsb_t(\cdot)$ is a function returning the t least significant bits of its argument.

   (b) Let $(b_1, \ldots, b_t)$ be the binary representation of the symbol to be transmitted. Leaf U_i computes the sequence $(d_1, \ldots, d_t)$ by doing:

   - If $(b_1, \ldots, b_t) = (c_1, \ldots, c_t)$ then $(d_1, \ldots, d_t) := (b_1, \ldots, b_t)$.
     Else $(d_1, \ldots, d_t) := (b_1 \oplus c_1, \ldots, b_t \oplus c_t)$

   (c) U_i generates a tn-bit sequence (where n is the number of leaves) I_i and sets the bits from the subsequence ranging from the $t(i-1)+1$

to the ti positions so that they match $(d_1, \ldots, d_t)$. The remaining bits are set to "0".

(d) U_i computes $\sigma_i := \mathcal{H}(v)^{\sum_{p=1}^{t} d_p \cdot x_{i,p}}$

(e) U_i sends the pair (I_i, σ_i) up to its parent node

3. MESSAGE AGGREGATION. An intermediate router R or the root S receives messages from its child routers/leaves and does the following.

(a) For each received pair (I_j, σ_j):

i. Let $i := 1$. Let $y := 1_G$.

ii. While $i \leq n$ loop

- $y := y \cdot \prod_{p=1}^{t} y_{i,p}^{I_j[t(i-1)+p]}$
- $i := i + 1$

iii. It checks $D(y, \mathcal{H}(v), \sigma_j)$. If this check fails, then $ERROR$.

(b) Once all expected messages $\{(I_j, \sigma_j)\}_j$ have been received, R aggregates them by computing $I = \bigvee_j I_j$ ($\vee$ denotes the bit-wise OR operation) and $\sigma = \prod_j \sigma_j$.

(c) If R is an intermediate node, it sends (I, σ) up to its parent node. Else, if it is the root, this is the final aggregated message.

4. SYMBOL EXTRACTION. From the final aggregated message (I, σ), the root obtains the symbol sent by each leaf as follows,

(a) Let $i := 1$

(b) While $i \leq n$ loop

- Compute $(c_1, \ldots, c_t) \leftarrow lsb_t(\mathcal{H}(v||K_i))$
- If $(I[t(i-1)+1], \ldots, I[ti]) = (c_1, \ldots, c_t)$ then $(b_{i,1}, \ldots, b_{i,t}) := (c_1, \ldots, c_t)$

- Else $(b_{i,1}, \ldots, b_{i,t}) := (I[t(i-1)+1] \oplus c_1, \ldots, I[ti] \oplus c_t)$
- $i := i + 1$

(c) Return $B = ((b_{1,1}, \ldots, b_{1,t}), \ldots, (b_{n,1}, \ldots, b_{n,t}))$, where $(b_{i,1}, \ldots, b_{i,t})$ is the binary representation of the symbol transmitted by U_i.

The security and cost analysis of this extension is not included since it would be done in the same manner described for the binary protocol. In this case, the message length tends asymptotically to tn.

Note that Step 2b above ensures that the sequence $(d_1, \ldots, d_n)$ does not have all its elements equal to 0. If this was the case, the signature σ_i would equal 1 and would lose its non-repudiation and integrity properties.

3.2 Secure many-to-one communications for resource-constrained devices

In Section 3.1, we presented [Sebe07b]. This protocol offered:

- Optimal message length.
- Lower computational cost than previous proposals in the literature.
- Immediate detection of corrupted data without requiring any extra error tracing procedure.

However, due to the use of GDH multisignatures, such protocol may not be adequate for resource-constrained devices.

In this section, we present a novel system for secure many-to-one symbol transmission which provides the same message length than [Sebe07b] but replaces the use of GDH cryptography with hash functions. This reduces the

computational cost at nodes. According to that, such a system is suitable for implementation in sensor networks. In addition to that, we also give an optimization of the proposed protocol to improve the efficiency of the message length and reduce the energy consumption at sensor nodes due to communication.

This system does not permit immediate detection of corrupted messages (like the previous scheme). Corrupted nodes are identified using an *a posteriori* tracing algorithm. Note that this tracing procedure is more efficient than the one presented in [Sebe07a].

3.2.1 General assumptions

Our protocol assumes a tree network where the root is the base station receiving data from the leaves (which may be sensor nodes). For the sake of simplicity, we assume that only the leaves send data. Intermediate nodes simply act as routers. Extending the proposed solution to accommodate data transmission from intermediate nodes is straightforward.

The base station (BS) is a full-fledged computer. Leaves and intermediate nodes are low-cost devices. The base station owns a private key SK_{BS}. The corresponding public key PK_{BS} is known and accepted as valid by all nodes in the tree. Let n be the number of leaves and U_i, $1 \leq i \leq n$, denote the leaves. Each leaf U_i shares a secret key K_i with the base station.

3.2.2 Many-to-one q-ary transmission

We represent each symbol from the q-ary alphabet by a different integer from the set $\{1, \ldots, q\}$. Parameter t is chosen as the smallest integer satisfying $q \leq 2^t - 1$. Parameter s is a security parameter (see Sections 3.2.4 and 3.2.5 for details about s). A protocol execution consists of the following steps:

1. CHALLENGE. The base station generates a random value v and signs it to obtain $\{v\}_{SK_{BS}}$. The signed value is multicast by the base station to all leaves.

2. MESSAGE GENERATION.

   (a) Upon receiving v and verifying its signature, each leaf U_i computes a pseudo-random t-bit sequence $(c_1, \ldots, c_t) \leftarrow lsb_t(\mathcal{H}(v||K_i))$, where $lsb_t(\cdot)$ is a function returning the t least significant bits of its argument, $\mathcal{H}$ is a one-way hash function and $||$ is the concatenation operator.

   (b) Each U_i computes a sequence $(d_1, \ldots, d_t)$ as follows. Let $(b_1, \ldots, b_t)$ be the binary representation of the q-ary symbol to be transmitted by U_i.

   - If $(b_1, \ldots, b_t) = (c_1, \ldots, c_t)$ then $(d_1, \ldots, d_t) := (b_1, \ldots, b_t)$
     Else $(d_1, \ldots, d_t) := (b_1 \oplus c_1, \ldots, b_t \oplus c_t)$

   Note that this step ensures that the sequence $(d_1, \ldots, d_t)$ does not have all its elements equal to 0. This all zeroes value is reserved to identify non-transmittal by leaves.

   (c) U_i computes an s-bit pseudo-random integer σ_i as follows:

$$\sigma_i \leftarrow lsb_s(\mathcal{H}(d_1, \ldots, d_t||v||K_i))$$

   (d) Each U_i generates a tn-bit sequence (n is the number of leaves) I_i and sets the bits from the subsequence between positions $t(i-1)+1$ and ti so that they match $(d_1, \ldots, d_t)$. The remaining bits are set to "0".

   (e) U_i sends the pair (I_i, σ_i) up to its parent node.

3. MESSAGE AGGREGATION. An intermediate node R (or the base station) receives messages from its child routers/leaves and does the following:

   (a) Store each received pair (I_j, σ_j) (they may have to be checked later).

   (b) Once all expected messages $\{(I_j, \sigma_j)\}_j$ have been received, aggregate them by computing $I = \bigvee_j I_j$ ($\vee$ denotes the bitwise OR operation) and $\sigma = \sum_j \sigma_j \pmod{2^s}$.

   (c) If R is not the base station, send (I, σ) up to its parent node. Else, this is the final aggregated message.

4. SYMBOL EXTRACTION. From the final aggregated message (I, σ), the base station obtains, for each leaf U_i, the binary representation $(b_{i,1}, \ldots, b_{i,t})$ of the symbol sent by the leaf. It is obtained from the sequence $(d_{i,1}, \ldots, d_{i,t})$, previously generated by U_i (see Step 2b), which is contained in I. Then the base station computes the pseudo-random integer linked to $(d_{i,1}, \ldots, d_{i,t})$ (see Step 2c), which will be used to check the integrity of the whole aggregated message. We next give the pseudo-code related to this process:

   (a) Let $i := 1$, $\omega := 0$.

   (b) While $i \leq n$ loop

   - Compute $(c_1, \ldots, c_t) \leftarrow lsb_t(\mathcal{H}(v||K_i))$.
   - If $(I[t(i-1)+1], \ldots, I[ti]) = (c_1, \ldots, c_t)$ then $(b_{i,1}, \ldots, b_{i,t}) := (c_1, \ldots, c_t)$.
   - Else $(b_{i,1}, \ldots, b_{i,t}) := (I[t(i-1)+1] \oplus c_1, \ldots, I[ti] \oplus c_t)$.
   - Compute $\phi_i \leftarrow lsb_s(\mathcal{H}(I[t(i-1)+1], \ldots, I[ti]||v||K_i))$.

- $\omega := \omega + \phi_i \pmod{2^s}$.
- $i := i + 1$.

(c) If $\omega = \sigma$ then return $B = ((b_{1,1}, \ldots, b_{1,t}), \ldots, (b_{n,1}, \ldots, b_{n,t}))$, where $(b_{i,1}, \ldots, b_{i,t})$ is the binary representation of the symbol transmitted by U_i. The base station also multicasts a signed acknowledgment $\{\text{“}Ack\text{”}||v\}_{SK_{BS}}$ to the leaves. This message contains the challenge v to avoid replay attacks. Upon receiving this message, intermediate routers remove messages stored at Step 3a. If $\omega \neq \sigma$ the base station launches the error-tracing procedure detailed in Section 3.2.3.

Figure 3.1 shows the message flow generated by a protocol execution in a simple scenario with a base station (BS), two intermediate nodes (R_1 and R_2) and four leaves ($U_1, \ldots, U_4$). In Figure 3.1.a the base station broadcasts a challenge to all leaves (Step 1 in the protocol execution). In Figure 3.1.b, message (1) sent by U_1 corresponds to the pair (I_1, σ_1) while message (2) sent by U_2 represents the pair (I_2, σ_2) (Step 2 in the protocol execution). Node R_1 constructs message (3), which corresponds to (I, σ), by aggregating messages (1) and (2) (Step 3 in the protocol execution). The same process occurs in the subtree rooted by R_2. The latter node constructs message (6) by aggregating messages (4) and (5), which correspond to the pairs (I_3, σ_3) and (I_4, σ_4), respectively. Eventually, the base station BS aggregates messages (3) and (6) to get the final aggregated message. After that, BS extracts the symbols transmitted by the leaves (Step 4 in the protocol execution).

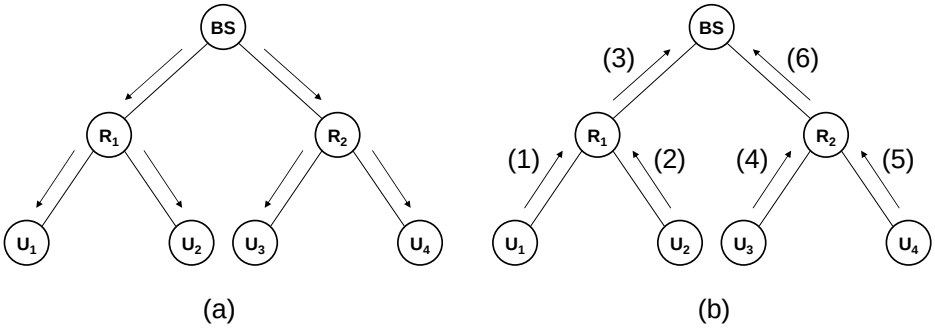


Figure 3.1: Message flow in a protocol execution

3.2.3 Procedure to deal with corrupted messages

During symbol extraction, the base station checks the integrity of the received message. If this verification fails, the base station identifies the message as corrupted. The following procedure allows to remove the corrupting nodes from the tree.

1. From the received I component, the base station computes the valid σ_i associated to each U_i:

   (a) For $i = 1$ to n do

   - $(d_{i,1}, \ldots, d_{i,t}) := (I[t(i-1)+1], \ldots, I[ti])$
   - $\sigma_i \leftarrow lsb_s(\mathcal{H}(d_{i,1}, \ldots, d_{i,t}||v||K_i))$

   (b) The base station sends to all nodes the signed message

$$\{I||\sigma_1, \ldots, \sigma_n||v\}_{SK_{BS}}$$

2. Upon receiving $\{I||\sigma_1, \ldots, \sigma_n||v\}_{SK_{BS}}$ and verifying its signature and the value v, each intermediate node R checks each stored message (I_j, σ_j) received from its children. For each (I_j, σ_j), R does:

(a) For each leaf U_i with nonzero contribution to I_j (that is, $(I[t(i-1)+1], \ldots, I[ti]) \neq (0, \ldots, 0)$), check that the contribution to I_j equals the contribution to I. If some of these checks fail, the point of corruption is above R's position and R stops the checking procedure.

(b) Else, R computes the sum modulo 2^s of the σ_i associated to each U_i who contributes to I_j. If the sum is equal to σ_j, the child who sent this message is considered innocent. If the result is different, that child is considered guilty.

When a malicious node has altered a message (note that a malicious node can corrupt a message at each protocol execution or only once in a while), all nodes located in the path from the corruption point to the root detect this corruption. However, the nodes detecting the corruption cannot decide whether corruption was caused by the child who sent the corrupted message or by another node located in the subtree rooted at this child. A simple solution would be to delete the entire suspicious subtree. This would entail the loss of a big part of the network due to a single corrupted message. We propose the following procedure to minimize the number of nodes to be eliminated:

1. The base station and all intermediate nodes have a pre-loaded integer λ associated to each child.

   - Initially, an intermediate node assigns a value $\lambda = 1$ to those of its children that are leaves.
   - If a child is an internal node, the initial assigned value corresponds to the number of leaves in the subtree rooted at that child.

2. When a node detects a corrupted message, it decrements the λ value assigned to the child from which this message comes from. Note that the λ values of all nodes located between the point of corruption and the base station will be decremented.

3. When the λ value assigned to a child becomes zero, the node closes communication with such a suspicious child (thus pruning the subtree rooted at this child).

The initial value assignment ensures that a corrupted leaf is pruned the first time it sends a bad message ($\lambda = 1$). The initial value assignment for an internal node is done assuming that it always acts honestly so that it only needs to be removed from the tree after all the leaves in its subtree have already been removed (in this case no more messages will come from its subtree). This procedure ensures that a dishonest internal node will be removed, although not necessarily the first time it corrupts and forwards a message.

3.2.4 Security analysis

We next explain the adversary model and the possible attacks the system has to be robust against. We refer to those attacks to justify the security properties achieved by our scheme: *confidentiality*, *authentication* and *integrity*. We also give the success probability of each possible attack.

Adversary model

Our attacker model considers an adversary who can control nodes (thus turning them into compromised nodes) and who can also access communication

Table 3.2: Possible attacks and their success probabilities

Attack	Success probability
Message eavesdropping	Not possible
Leaf impersonation	$1/((2^t - 1)2^s)$
Leaf contribution alteration	$1/2^s$
Leaf contribution removal	Not possible
Message aggregation disruption	Not possible

lines to capture, modify and retransmit messages. The attacker's computational power does not permit her to break current computationally secure cryptosystems.

We consider that an adversary can try the following attacks:

- Eavesdrop messages.
- Impersonate a certain leaf U_i.
- Alter the contribution of a certain leaf U_i.
- Remove the contribution of a certain leaf U_i.
- Disrupt the aggregation of messages received from some child nodes.

More specifically, an external attacker can try the first four attacks while an adversary who has compromised some nodes could also try the last one.

Attacks and security properties

Table 3.2 summarizes the success probabilities of each possible attack in scenarios where non-transmittal is disallowed. We next justify the values in the table.

Message eavesdropping. This attack refers to the *confidentiality* property. An adversary eavesdropping messages of the form (I, σ) at some point (called sniffing point from now on) in the communication tree can discover, by observing the bits ranging from $I[t(i-1)+1]$ to $I[ti]$, the sequence $(d_{i,1}, \ldots, d_{i,t})$ transmitted by each leaf U_i located below the sniffing point. The attacker will be unable to decrypt this sequence because decryption requires knowledge of the secret key K_i.

The σ value does not provide any useful information to the attacker either.

From I, an adversary can determine which leaves transmitted and which ones did not. In applications where this fact causes information disclosure, non-transmittal should not be allowed.

Leaf impersonation. This attack refers to the *authentication* property. An intruder (who does not know K_i) trying to send a given symbol coming from U_i faces several difficulties.

First of all, the q-ary symbol is encrypted prior to encoding it inside I_i. Since the attacker does not know the encryption key, she can only fill the corresponding t bits in I_i randomly. The probability that decryption of those bits leads to the desired q-ary symbol is $1/(2^t - 1)$.

On the other side, the redundancy σ_i is also computed using K_i. In this way, the probability of randomly guessing the appropriate σ_i is $1/2^s$.

Leaf contribution alteration. This attack refers to the *authentication* and *integrity* properties. In case the attacker captures and alters the contribution of a leaf, the difficulty comes from the low probability of guessing σ, which is $1/2^s$. A sufficiently large value s exponentially reduces the chances of a corrupting attacker to stay undetected.

Leaf contribution removal. The contribution of U_i can be easily erased if σ_i is known. This fact will be considered as a non-transmittal by the base station. To detect these erasure attacks, non-transmittal should be disallowed.

In case the adversary does not know the σ_i value generated by U_i, she must guess the appropriate σ_i with a probability of success $1/2^s$.

Disruption of the aggregation of child node messages. A dishonest intermediate node can decide not to aggregate a message received from some of its child nodes. The base station will consider this fact as a non-transmittal. To detect this situation non-transmittal should be disallowed.

3.2.5 Performance analysis

We next evaluate the protocol performance in terms of message length and computational cost at the sensor nodes. We also give an optimization of the proposed protocol to improve the efficiency of the message length and reduce the energy consumption at nodes due to communication.

Table 3.3 summarizes the performance results obtained. The values in the table are justified below.

Message length

Our protocol keeps the length of messages constant on their way from the leaves towards the base station. Each message consists of the pair (I, σ). Component I encodes the q-ary symbol transmitted by each leaf and its bitlength is tn, where n is the number of leaves of the multicast tree and $q \leq 2^t - 1$. Since t is a constant, the bitlength of I is $O(n)$. Component σ has

Table 3.3: Performance results

Item	Cost
Message length (for $n \uparrow\uparrow$)	tn
Message length in the error-tracing procedure (for $n \uparrow\uparrow$)	$(t+s)n$
Cost of message generation	$O(n)$
Cost of message aggregation	$O(n^2)$
Cost of symbol extraction	$O(n)$
Cost of error tracing at intermediate nodes	$O(n^2)$

a constant length of s bits. So its length is $O(1)$. In this way, messages have a bitlength $O(n)+O(1)$. For large values of n, this length asymptotically tends to tn. Note that this length is linear, which represents a limitation on the total amount of leaves which can participate in the network. However, it can be proven that symbols sent by n leaves cannot be aggregated in a message whose length is below $O(n)$ when all symbols have the same probability of being sent.

In the event of a corrupted message, the base station multicasts a message which contains $(I||\sigma_1, \ldots, \sigma_n||v)$. The first component has tn bits, the second one sn bits and the last one $O(1)$ bits. Thus, this special message has a bitlength of $O(n)$. For large values of n, this length asymptotically tends to $(t+s)n$.

Computational cost

Next, we analyze the time complexity of the protocol in four operations: message generation, message aggregation, symbol extraction and error tracing at intermediate nodes.

Message generation. Each message has two components (I_i, σ_i). Leaf U_i employs $O(n)$ time to generate the binary sequence I_i. The computation

of σ_i does not depend on n, so it is $O(1)$. During message generation, each leaf verifies one signature and computes two hash functions and at most t bitwise XOR operations.

As mentioned in Section 1.1.1, there are fast algorithms for digital signature verification in resource-constrained environments. As a real example, the authors of [BlaB05] implement the Elliptic Curve Digital Signature Algorithm [Ecds99, John01] on a MICA2 mote [Berk04, Mica08], designed by researchers at the University of California at Berkeley. This device offers an 8-bit, 7.3828-MHz ATmega 128L processor, 4 kilobytes (KB) of primary memory (SRAM), and 128 KB of program space (ROM). According to their results, an ECDSA signature verification in such a device takes about 24.17 seconds (this time can only be expected to decrease as technology progresses).

The remaining operations (hash functions and bitwise operations) take negligible time. Thus, our scheme is suitable for resource-constrained leaves.

Message aggregation. An intermediate node receives and aggregates messages. Aggregation of $\{(I_j, \sigma_j)\}_j$ into (I, σ) requires at most $O(n)$ bitwise OR operations over $O(n)$-long messages during the computation of I. Computation of the new σ requires at most $O(n)$ additions modulo 2^s (each with cost $O(1)$). This results in a maximum $O(n^2)$ cost. Note that intermediate nodes compute only bitwise operations and additions modulo 2^s. These operations are appropriate for resource-constrained nodes.

Symbol extraction. The base station extracts the contribution of each leaf

in a vector $B = ((b_{1,1}, \ldots, b_{1,t}), \ldots, (b_{n,1}, \ldots, b_{n,t}))$ by processing component I (total length tn bits). Since t is a constant, this time is $O(n)$. Integrity checking does not increase this cost.

Since the base station is a full-fledged device, we consider that all operations executed in this step are affordable.

Error tracing at intermediate nodes. This procedure is only invoked in case of a corrupted message event. An intermediate node may need to verify a signature and check the value I_j sent by each of its children. Each check takes $O(n)$ time. Since the number of children may also be $O(n)$, the maximum time spent on this operation is $O(n^2)$. Checking σ_j has at most the same cost.

In this step we require one signature verification. As explained above, this operation is affordable on real sensor nodes like the MICA2 mote. In addition to that, intermediate nodes must execute $O(n^2)$ additions modulo 2^s. These operations are suitable for resource-constrained nodes too.

Message length optimization

Our system is designed for nodes that are resource and power-constrained devices. This motivates the need to reduce energy consumption as much as possible. Reducing the length of messages is one way to achieve this.

In our protocol, leaf U_i sends (I_i, σ_i) where I_i is a tn bit long binary sequence. Useful information within I_i is contained in bits located between positions $t(i-1)+1$ and ti. The remaining bits of I_i are set to 0.

This information could be represented in a more compact way using $\log n$ bits to code index i and t bits for useful information. In this way, the length

of I_i would be $t + \log n$. Aggregation of vectors I_i would be done by concatenation. In this way, the length of a vector I containing data from j leaves would be $j(t + \log n)$ bits.

For small values of j this results in shorter messages than those described in our protocol above (*i.e.* when $j(t + \log n) < tn$). Low values of j appear at nodes that are far from the root. However, when j grows towards n this new coding results in longer messages than those described above.

Therefore using this alternative coding when j satisfies $j(t + \log n) < tn$ (near the leaves) and switching to the initial coding when messages get near the root is a way to minimize the length of transmitted data.

Chapter 4

Secure and private information sharing in MANETs

In this chapter we present two contributions related to information sharing in mobile ad hoc networks. Both contributions provide security and privacy for the users of the network. Incentives are given to thwart user misbehavior.

Section 4.1 presents an information system aiming to provide information *just in time* and *just in place* in a specific area. This system is deployed in a city within which a user, regardless of her location, can request information any time using her mobile device and its wireless connection. Access to information is made possible by a metropolitan ad hoc network based on peers enabled with wireless technology. A typical application of this system would be tourist information: a person touring a city can query the system to obtain a list of museums near her current location or some information about a given historical building she is currently visiting. This proposal has been published in [Cast07].

Section 4.2 presents a new scheme designed to disseminate advertisements through mobile ad hoc networks. This proposal exploits the capabilities

of mobile ad hoc networks to increase the visibility of the products being offered by merchants. The starting point is a merchant who generates an advertisement that is subsequently disseminated by citizens who carry mobile devices acting as network nodes. This scheme has been published in [Viej07].

4.1 Distributed information provision

In this section, we present an architecture for mobile ad-hoc networks which has been published in [Cast07]. This scheme offers distributed information provision in urban environments. It requires some collaborative users to become information providers. Such volunteers are rewarded for their work. The proposed architecture is specified as a protocol suite taking security and privacy aspects into account.

4.1.1 System overview

We focus on an environment where some end-user nodes build a wireless ad hoc network and act as information providers. These nodes devote some of their computational resources (storage, bandwidth, processing power) to storing and serving information. In this way, when another user in the network requests some information, those nodes storing the requested information can supply it.

Volunteering to become an information provider is rewarded depending to the amount of served information requests. Every time a server provides information to some user, it obtains a receipt allowing it to prove that it has performed the service. Periodically, the server contacts the main content provider (*i.e.* the main source of content, in our case study the city tourist office) to get paid according to the number of requests it has served. This is a

way to encourage users to devote more resources to provide information. The more information a server stores, the more requests it will be able to serve, and thus, the more money it will receive for its service. A server located at a certain place will probably contain information that may interest nearby users.

4.1.2 System architecture

In this section we present the system components: entities, messages exchanged between entities and protocols between entities.

Entities

- *Content Source (CS).* This is the entity offering the information service. In the aforementioned example about tourist information, this entity may be the tourist office of a city holding information of particular interest for residents or visitors. Some examples could be:
  - Information on historical landmarks, including short multimedia videos, audio streams and digital documents on them.
  - Schedule of cultural and leisure activities, like cinema or theater, including trailer viewing options.
  - Information about restaurants: opening hours, menus, prices.
  - Location of services: police stations, hospitals, pharmacies.
- *Users.* Users whose devices form the ad hoc network. We distinguish two kinds of users:
  - Those that query the system when they need information. Normally, they use a mobile device and request information through

the ad hoc network. We refer to them as *end-users (EU)*.

– Those that devote part of their computational resources to storing and serving some of the information supplied by the content source. These users not necessarily use a mobile device. They could store information in their desktop PC with an ad hoc network interface. We refer to these users as *server-users (SU)*.

Messages

We distinguish two types of communication:

- The first type follows a client-server paradigm and involves the content source CS. We have chosen this approach because the communication between CS and the other two entities (SU, EU) only occurs at very specific moments and is unlikely to cause a bottleneck.
- The second type of communication, the dialog between an EU and a SU, follows a peer-to-peer (P2P) paradigm.

Messages consist of two parts. The first part contains the message itself, divided in two or three sections: message type, sender identifier (in the P2P environment) and message body. The second part contains the cryptographic data: the signature on the first part of the message, the algorithm used to calculate the signature and the digital certificate for the sender's public key. This structure allows the receiver to verify the validity of the message.

Protocols

We use the following notations to describe our protocols:

- P_{entity}, S_{entity}: Asymmetric key pair of $entity$, where P_{entity} is the public key and S_{entity} is the private key.

- $S_{entity}[m]$: Digital signature of message m by *entity*. By digital signature we refer to computing the hash value of message m using a collision-free one-way hash function and encrypting this hash value using the private key of *entity*.
- $E_{entity}(m)$: Encryption of message m under the public key of *entity*.
- $D_{entity}(c)$: Decryption of message c under the private key of *entity*.
- $H(m)$: Hash value of message m using a collision-free one-way hash function.
- $m_1||m_2$: Concatenation of messages m_1 and m_2.

We next detail the different protocols used by the entities participating in the system:

End-user registration. To register as an end-user, a candidate user must contact the CS and install the necessary application software. *e.g.* in our tourist information case study we assume that there exist several places in the city (*e.g.* airport, railway station or tourist office) where a user can register.

The end-user registration protocol is as follows:

Protocol 1

1. *The user does:*
   (a) *Obtain the following information from the CS:*
      - *Internet address from which to download the application software.*

- *Validity period,* i.e. *time window during which the user will be allowed to use the system.*
- *Access code to download and install the software.*

(b) *Connect her device to the Internet and download the application software.*

(c) *Install the application software.*

(d) *Run Procedure 1 below and obtain the private key S_{EU} in a PKCS#8 file [Pkcs08], and a Certificate Signing Request (CSR).*

(e) *Send the CSR to the Content Provider.*

2. *The CS does:*

(a) *Issue the user's certificate using the CSR.*

(b) *Add the issued certificate to the CS database.*

(c) *Send the issued certificate to the user.*

3. *The user stores the following information in a PKCS#12 [Pkcs08] file:*

- *User private key S_{EU}.*
- *User certificate.*
- *CS certificate.*

Procedure 1

1. *Generate a private/public RSA key pair [Rsa78].*
2. *Store the private key in a PKCS#8 file.*
3. *Generate a Certificate Signing Request (CSR). The file must use the PKCS#10 [Pkcs08] standard.*

4. *Return the PKCS#8 file and the CSR.*

Server-user registration. A user wishing to register as a server-user contacts the CS from whom she will receive a unique identifier and the software that will enable her to serve information. Afterwards, the user generates a private/public key pair and sends her public key and her identifier to the CS in order to get the corresponding certificate. Finally, the user indicates the desired information items and downloads them to her hard disk.

More formally, the server-user registration protocol is as follows:

Protocol 2

1. *The user does:*
   (a) *Sign a contract with the CS specifying the user's rights and duties.*
   (b) *Send the user's bank data for future payments to CS. For confidentiality, these data are sent encrypted under the public key of CS.*
2. *CS does:*
   (a) *Generate a unique identifier Id.*
   (b) *Send to the user the unique identifier Id and the software that will enable the user to serve information. For confidentiality, Id is sent encrypted under the public key P_{EU} (the candidate server-user is assumed to be already an end-user with a private/public key pair (S_{EU}, P_{EU})).*
3. *The user does:*

   (a) *Run Procedure 1 to obtain the private key S_U in a PKCS#8 file, and a Certificate Signing Request (CSR).*

   (b) *Send the CSR to CS.*

4. *CS does:*

   (a) *Issue the user's certificate using the CSR.*

   (b) *Add the issued certificate to CS's database.*

   (c) *Send the issued certificate to the user.*

   (d) *Send the catalog information.*

5. *The user stores the following information in a PKCS#12 file:*

   - *User private key S_U.*
   - *User certificate.*
   - *CS certificate.*

Information request. When an end-user requests an information item, the query reaches several server-users. Among these, those holding the requested item return a positive acknowledgment. Then, the end-user downloads the requested information from a particular server-user selected among those which have sent positive acknowledgment. Finally, the end-user sends a receipt to the selected server-user. As we will see later on, the SU will use this receipt in order to claim the corresponding reward from the CS.

Formally, the information request protocol is as follows:

Protocol 3

1. *The end-user EU computes a request in order to obtain a specific information, where the request consists of the following data:*

- *Description of the requested item, I.*
- *Date and time of the request, T_r.*
- *Digital signature of I and T_r, $S1 = S_{EU}[I||T_r]$.*

This query spreads using a broadcast approach. Therefore, all SU who are close to EU receive her request.

2. *Each server-user SU who receives the query does:*
   (a) *Verify the digital signature $S1$ using EU's public key.*
   (b) *Search for the information.*
   (c) *If I is in SU's database, reply to EU. The reply contains the following data:*
      - *User's request, $R_{EU} = I||T_r$.*
      - *Date and time of the answer, T_a.*
      - *Digital signature on R_{EU} and T_a, that is, $S2 = S_{SU}[R_{EU}||T]$.*
3. *EU does:*
   (a) *Collect the replies from the SUs. Without loss of generality, assume that the set of SU replying to EU is $SU_1, SU_2, \cdots, SU_n$. See Section 4.1.4 below on the value of n.*
   (b) *Verify the digital signatures of the SUs, that is, $S2_1, S2_2, S2_3, \cdots, S2_n$ using the public keys of each SU.*
   (c) *Choose one server-user SU' $\in \{SU_1, SU_2, \cdots, SU_n\}$. This choice can be performed in a way to maximize privacy (see Section 4.1.4 below).*
   (d) *Send a request to SU' with the following data:*
      - *Description of the requested information, I.*

- *Date and time of the request,* T_r.
- *Identifier of the node this request is addressed to,* $Id_{SU'}$.
- *Digital signature on* I, T_r *and* $Id_{SU'}$, *that is,* $S3 = S_{EU}[I||T_r||Id_{SU'}]$.

4. *SU' does:*
   (a) *Verify the digital signature* $S3$ *using the public key* P_{EU}.
   (b) *Send the following message:*
   - *Description of the requested information,* I.
   - *Requested information,* $Info$.
   - *Date and time of the answer,* T_a.
   - *Digital signature of the* I, $Info$ *and* T_a, *that is,* $S4 = S_{SU'}[I||Info||T_a]$.

5. *EU does:*
   (a) *Verify the digital signature* $S4$ *using* P_{SU}.
   (b) *Check whether the received data correspond to the information requested.*
   (c) *If the check is OK, issue a receipt and send it to SU' with the following data:*
   - *Description where EU asserts that she has received the item described as* I *from SU'.*
   - *Date and time,* T.
   - *Identifier of SU',* $Id_{SU'}$.
   - *Digital signature on* I, T, *and* Id_{SU}, *that is,* $S5 = S_{EU}[I||T||Id_{SU}]$.

6. *SU' does:*
   - *Receive the receipt.*
   - *Verify* $S5$ *using* P_{EU}

- *Store the receipt.*

Server-user payment. As previously described, server-users get a receipt every time they serve information. These receipts are stored. Once a large enough batch of receipts has been collected, a server-user contacts CS to get paid for the services provided. Note that sending receipts one at a time to CS would be very inefficient. The reason is that, since the reward for a single service is very low, the processing costs of such a payment would be too significant.

The protocol to redeem a batch of receipts is as follows:

Protocol 4

1. *SU sends the receipts to CS.*
2. *CS does:*
   (a) *Verify the digital signature of each receipt*
   (b) *Check for duplicated receipts*
   (c) *Compute the money that must be paid to the information node*
   (d) *Transfer the money to the bank account of SU*

4.1.3 Security analysis

Our communication protocols use different types of messages to be transmitted in each phase. Every exchanged message contains a plaintext part and a valid signature. The plaintext part contains the information transmitted between nodes and the signature provides *authentication*, *integrity* and *non-repudiation* to such messages.

Confidentiality

In principle, confidentiality is only implemented in the server-user registration protocol, when the user sends her bank data to CS and CS returns a unique identifier (Steps 1 and 2 of Protocol 2). The rest of messages are assumed to be non-confidential, which is plausible for most applications (*e.g.* tourist information). However, if confidentiality is required, it can be achieved by encrypting messages under the public key of the intended receiver.

Collusion security

Collusion between end-users and server-users to obtain unlawful rewards is conceivable: some end-users perform a huge amount of information requests to certain server-users, and the latter then share with the former the rewards obtained from the CS.

A possible solution is to charge the end-users a small fee for enjoying the information service. This payment can be performed using offline electronic checks as stated in [Chau90] or any micropayment system (*e.g.* PayWord, [Rive96]).

However, one must acknowledge that collecting payment from the end-users can jeopardize the success of many applications, like the tourist information system. Therefore, a preferred countermeasure against user collusion is for the CS to record and analyze the number of receipts submitted by the SUs and the number of receipts issued by the same EU. Since each receipt contains the exact time and date when it was issued, a limit on the number of requests that an EU can perform within a period of time can easily be enforced. The CS will not honor any receipts beyond those that can be issued by a certain user; furthermore, as soon as CS detects that an end-user

has issued more receipts than allowed, CS alerts the SUs to stop serving any further request from that suspect SU. The SUs receive this alert when they synchronize resources with the CS or when they redeem their receipts. In this way, the effects of possible user collusions are tolerably mitigated.

4.1.4 Privacy analysis

In any information service, end-user profiling is a real threat. Indeed, information providers can keep track of the requests submitted by end-users, with a view to investigating their tastes, preferences, locations, etc. This is clearly a potential privacy violation.

In a conventional information service where end-users get information directly from a single information provider, one often assumes that information provider to be trusted or at least not to be interested in violating the privacy of end-users. At any rate, if there ever were any provable violation, the information provider would be liable and could be charged accordingly.

In a peer-to-peer mobile ad hoc information service, the privacy problem is much more serious. End-users obtain information through server-users who are occasional information relayers and cannot be trusted to the same extent as to privacy preservation.

End-user privacy can be significantly increased by using an alias when registering as an end-user and by properly tuning Protocol 3:

- When the end-user application detects that there are server-users among the n replying to Step 3a who already replied to more than p requests from the same end-user in the past (p is a privacy parameter), the application warns the end-user of a potential privacy problem. The end-user has two choices: either move to a different area where she will find different server-users or to go ahead and jeopardize her privacy.

- In Step 3c, a wise policy is for the end-user application to choose the server-user which has replied to least requests to the end-user in the past.

Of course, we are assuming that the server-user application has not been tampered with, so that: i) it replies when the server-user hears a request for an information item it holds; ii) it forgets about requests for information items the server-user does not hold.

In the presence of malicious server-users, a combination of the following two strategies can be useful:

- Use short validity periods for end-users, which will force end-users to frequently re-register under a new alias.

- Avoid issuing many information request from the same place, which should be easy for a roaming end-user (*e.g.* tourist visiting a city). Moving to another area is a way to get rid from the current server-users, both the legitimate and the malicious ones.

4.2 Advertisement dissemination

In this section, we propose an advertisement dissemination model which has been published in [Viej07]. This scheme exploits the capabilities of mobile ad hoc networks to increase the visibility of the products being offered by merchants. It offers incentives to stimulate the collaboration of nodes. Cryptographic techniques are used to prevent manipulation and preserve the privacy of users. Specifically, the AdPASS [Stra04] system is outperformed in the following aspects:

- Security is achieved against (individual or colluding) dishonest nodes trying to modify transmitted advertisements in order to unlawfully increase their share of incentives.

- Privacy is preserved without resorting to any trusted third party. Our system only requires a certification authority (CA) to certify the merchant's public key. In any case, this authority can not disclose users' identities.

- The incentives rewarding a certain purchase are distributed among all co-operating users given on how long they have held the advertisement leading to that purchase before transferring it to another user. This is a fair proposal which does not restrict the advertisement's dissemination range.

Our scheme uses multisignatures over a Gap Diffie-Hellman group [Bold03]. The required mathematical background has been previously introduced in Section 3.1.1.

4.2.1 System overview

Our protocol assumes the existence of a merchant and several mobile nodes that communicate through a MANET. We assume the existence of dishonest users (who may act individually or in collusion) interested in obtaining a higher reward than the one they are entitled to. We do not require the users to be registered with any central entity. Thus, our system is appropriate for very dynamic environments where connectivity to a central entity may not be guaranteed.

Functionally speaking, a user holding an advertisement actively contacts users within her range and sends them the content of the advertisement.

Initially, the advertisement is held by the merchant. Some of the contacted nodes may purchase the advertised good and/or be interested in holding the advertisement themselves for further dissemination.

On the occasion of a purchase request, the buyer sends to the merchant the advertisement (if any) which has motivated her purchase; attaching the advertisement entitles the buyer to a discount. The incentives rewarding that purchase are distributed among the nodes in the path from the merchant to the buyer proportionally to the time they have held the advertisement. E-coins are used to pay those incentives.

In order to facilitate the distribution of incentives, when an advertisement is transferred to a new holder, a time stamp indicating the moment of the transfer is added to the advertisement. In this way, when an advertisement comes back to the merchant together with a purchase request, the merchant can ascertain the incentive that corresponds to each collaborating node. The system is totally anonymous, *i.e.*, the information that nodes add to an advertisement does not allow to identify them. Also, different contributions of a node to different advertisements cannot be related. In this way, unlinkability is also provided. Obviously, we are assuming that the appropriate measures are being taken to avoid node tracking by other means (for instance, frequent change of MAC and IP addresses).

The above system is sustainable for the merchant, who never loses money, because incentives are only paid for advertisements which generated a purchase.

4.2.2 Set of protocols for advertisement dissemination

We next describe the six protocols that conform the proposed system.

Advertisement generation

Merchant M has its public key, PK_M, and its digital certificate issued by a Certification Authority, $Cert_{Aut}\{PK_M\}$. We denote by SK_M the secret key corresponding to PK_M.

1. When M wants to promote a product, it generates an advertisement α containing its public key certificate, the offer description and the expiration time of this offer:

$$\alpha = \{Cert_{Aut}\{PK_M\},\ Description,\ ExpirationTime\}$$

   This advertisement is signed by M to obtain $\{\alpha\}_{SK_M}$.

2. A node U_i interested in disseminating the advertisement contacts M and receives the following message:

$$\beta = \{\alpha,\ PubKeyChain,\ Multisignature,\ TimeChain\}$$

   The fields of β are initialized as follows:

   - $PubKeyChain$ is an ordered list initially left empty;
   - $Multisignature$ is initialized to $\{\alpha\}_{SK_M}$;
   - $TimeChain$ is an ordered list initially containing a single element that is a tuple formed by $Time$ and its signature $\{Time\}_{SK_M}$; $Time$ corresponds to the time this operation has been performed.

3. U_i checks β (see the protocol for advertisement checking). If all checks are correct, U_i accepts the advertisement from M and starts its dissemination.

Advertisement dissemination

Upon accepting an advertisement, U_i informs other nodes about the offer it contains. Due to the inherent mobility in the nodes, U_i is likely to disseminate the offer quite far from M.

Additionally, when U_i contacts a nearby node U_j, U_i asks whether U_j is interested in disseminating the advertisement (our scheme is not linked to any specific framework to perform such initial contact between users, the one presented in AdPASS [Stra04] can be used). If she is, they will start the advertisement transfer. In order to guarantee anonymity and unlinkability, nodes must change their MAC and IP addresses after each contact.

Note that, after an advertisement transfer from U_i to U_j, U_i still holds the advertisement and can continue its dissemination and transfer to other nodes. In this way, the number of nodes disseminating a certain advertisement can grow exponentially.

Advertisement transfer

The advertisement transfer protocol requires users U to have a public/private key pair (PK_U/SK_U). To provide unlinkability, this key pair has to be changed after each execution. Before renewing her key pair, a user stores the secret key. This key will be needed in order to receive the incentives (as will be detailed next in the incentive payment protocol).

1. A user U_j interested in an advertisement α held by another user U_i asks U_i to transfer it.

2. U_i appends her public key to the value $PubKeyChain$ in β. This is

$$PubKeyChain' := PubKeyChain \cup PK_{U_i}$$

3. U_i Computes the signature $sig := \{\alpha\}_{SK_{U_i}}$. Then she computes

$$Multisignature' := Multisignature \cdot sig$$

4. U_i obtains the current time, signs it and appends the signed time to the time chain, that is: $TimeChain' := TimeChain \cup \{Time \mid\mid \{Time\}_{SK_{U_i}}\}$ (at the end).

5. U_i generates

$$\beta' := \{\alpha,\ PubKeyChain',\ Multisignature', TimeChain'\}$$

and sends it to U_j.

6. U_i stores the secret key SK_{U_i} and generates a new key pair that will be used at the next transfer.

7. U_j checks β' (see the protocol for advertisement checking). If all checks are correct, U_j informs other nodes about the offer in β'.

Advertisement checking

A user U_i receiving a message β should check its validity prior to accepting it. This is done as follows:

1. Check the validity of $Cert_{Aut}\{PK_M\}$ (obtained from α). This requires checking the signature by the authority, its expiration date and, if possible, its revocation status.

2. Compute the product of all public keys contained in $PubKeyChain$ and PK_M. Let us denote by $GlobalKey$ the result of this operation.

3. Check that $Multisignature$ is a correct signature over α that is validated using $GlobalKey$.

4. Check that $ExpirationTime$ (obtained from α) has not expired.

5. Check that the first element of $TimeChain$ is a correct signature generated by the Merchant.

6. For each key contained in $PubKeyChain$, check that the j-th public key in $PubKeyChain$ can validate the $(j+1)$-th signature in $TimeChain$.

7. Finally, check that the values of elements in $TimeChain$ are sorted in ascending order and that the last element corresponds to the current time.

Advertisement deposit

A user U_i interested in the product advertised in β contacts the merchant and buys it. By sending β to the merchant, U_i will obtain the price reduction detailed in β. This price reduction motivates users to deposit advertisements.

Incentive payment

Once a merchant sells a product to a customer who has deposited an advertisement, it has to pay the incentives to all users who have collaborated in its dissemination.

The merchant gives a fixed amount of money for each received advertisement. This amount of money is divided between collaborating nodes proportionally to the time each collaborating node has held the advertisement along the path from the merchant to the buyer (see Section 4.2.4 for details about the model used to reward incentives). This information can be

obtained from the values in $TimeChain$. The merchant does not know the identity of the nodes that collaborated in the advertisement distribution. It only knows their public key. For each payment the merchant authorizes her bank to issue an e-coin. Let us assume user U_i (who remains anonymous and is only known by her public key) has to receive an e-coin for a given value v.

The merchant sends a message to her bank indicating that she can issue an e-coin with value v to any person providing password p. Then, the merchant publishes a message in a public repository containing p encrypted with the public key of U_i. This indirect procedure through a public repository is needed because U_i is anonymous and may be temporarily out of range.

Later, U_i checks the repository, decrypts the message and obtains p. Using this password, the bank permits her to obtain an e-coin (through the corresponding e-coin issuing protocol). The e-coin system must be anonymous such as the one proposed by Chaum in [Chau89]. This is because the e-coin may later be spent non-anonymously (for instance, if the purchased product has to be delivered by courier). If the e-coin system was not anonymous, it could be possible to link the identity of the person spending the e-coin to the public key used in the dissemination protocol.

4.2.3 Example of an advertisement dissemination

We next clarify the operation of this scheme following the communication steps described above. We base our explanation on the graphical example shown in Figure 4.1.

1. Advertisement generation. The merchant wants to promote a certain product and generates an advertisement and informs about it

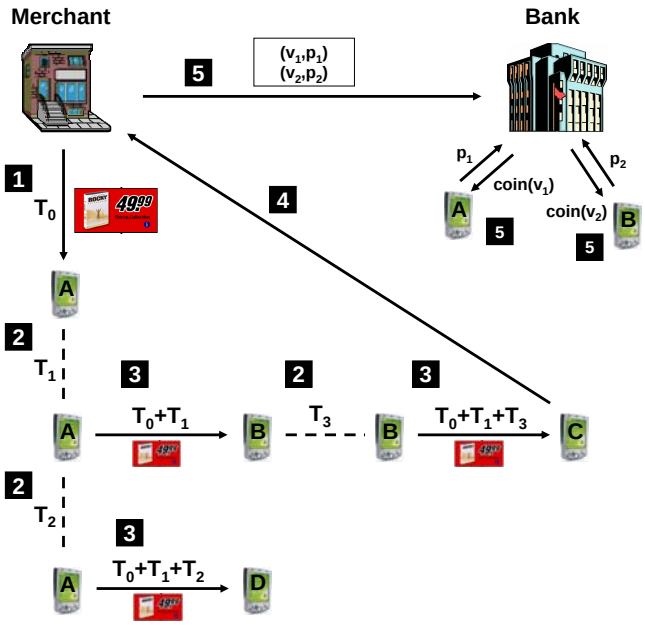


Figure 4.1: Graphical example of an advertisement dissemination

the users within range. User A is interested in disseminating this advertisement and contacts the merchant to request transfer of the advertisement β. Then A checks the validity of β and starts its dissemination. This occurs at time T_0.

2. Advertisement dissemination. A roams around while informing other nodes she meets about advertisement β. Then, A transfers the advertisement to two interested nodes B and D at times $T_0 + T_1$ and $T_0 + T_1 + T_2$ respectively. At time $T_0 + T_1 + T_3$, node B transfers the advertisement to node C.

3. Advertisement transfer. In each transfer, the node which receives the advertisement checks its correctness (see the protocol for advertisement checking) prior to accepting it.

4. Advertisement deposit. User C is interested in the product advertised in β. Therefore, she contacts the merchant and buys it. By sending β to the merchant, C will benefit from the price reduction detailed in the offer.

5. Incentive payment. The merchant uses the values in the $TimeChain$ embedded in β to determine that A has carried this advertisement during time $T_0 + T_1$ and B has carried it during T_3. Then, the merchant sends a message to its bank indicating that it can issue two e-coins for values $v_1(T_0 + T_1)$ and $v_2(T_3)$ to any person providing passwords p_1 and p_2 respectively. The joint value of those two e-coins is the fixed amount that the merchant is willing to pay for each completed sale of the product. Finally, the merchant publishes p_1 and p_2 encrypted with the public key of A and B respectively in a public repository.

   Later, A and B check the repository and obtain their respective password. Then, they contact the bank and obtain their respective e-coin through the corresponding e-coin issuing protocol.

4.2.4 Comparison to other reward models

As explained before, in our scheme the merchant divides a fixed amount of money between the nodes which have collaborated in an advertisement dissemination. The money earned by a certain node is proportional to the time which such a collaborating node has held the advertisement along the path from the merchant to the buyer. We next explain the advantages of this approach in comparison with the model presented in [Stra04] and with a simple model where each node receives money each time it collaborates (this scheme does not consider how long a node has held the advertisement, only

if the node has held it or not).

In [Stra04], the merchant fixes an amount of points as reward to a certain advertisement. Each user who collaborates in the dissemination will claim the number of points that she desires. This means that if a greedy user U_i claims too many points, the advertisement will not be disseminated by any other user since there will not be enough remaining points. Thus, this represents a strong restriction in the advertisement's dissemination range. Besides, users are not rewarded in a fair way and this motivates the users to apply strategies for keeping and passing along points instead of collaborating in the dissemination.

The simple model is fairer than [Stra04]. Each U_i which takes part in a dissemination will receive the same amount of money. However it has two main problems:

1. If there is no limit in the number of hops, there is no limit either in the amount of money that the merchant must give as incentives. This represents a major concern for the merchant. We can solve this problem by enforcing an upper limit but then the advertisement dissemination range will be restricted like in [Stra04].

2. Since the merchant gives incentives to each user who collaborates, a certain user with n identities can transfer a certain advertisement to herself $n-1$ times (using her $n-1$ alternative identities). At the end of the process, this user will get incentives for each of her n identities.

To solve these two problems we propose to add a second dimension (how long a user holds an advertisement) to the simple model. Besides, the merchant establishes a fixed amount of money (incentives) that will be divided between the collaborating users. We next explain how our proposal affects

the two problems stated:

1. The merchant after each sale divides the money assigned to pay advertisement dissemination between collaborating nodes proportionally to the time each collaborating node has held the advertisement. It means that the merchant never loses money. Besides, users will always receive incentives, although a node which has held a certain advertisement for a short time in comparison with others will probably get a very small amount of money.

2. A certain user which holds an advertisement for n epoch (interval of time) will get the same amount of money than a dishonest user which has n different identities and holds the advertisement for one epoch with each identity.

4.2.5 Security and privacy analysis

We next explain the adversary model and the possible attacks the system has to be robust against. We refer to such attacks to prove the security properties achieved by our scheme: integrity, authentication and non-repudiation. We also explain how privacy (anonymity and unlinkability) is obtained.

Adversary model

In our system, an adversary is any entity or group of entities wishing to disrupt normal system operation or aiming to collect information on nodes who have collaborated in advertisement dissemination. The nodes that can take part in a dishonest coalition are:

- *The merchant.* It may wish to identify and/or trace nodes who collaborate by spreading announcements. It may also repudiate having

generated a certain offer.

- *The bank.* It may wish to identify and/or trace nodes who collaborate in message dissemination.
- *Dishonest users.* They may wish to alter advertisements so as to increase the amount of their assigned reward. They may also wish to inject false disrupting data or identify and/or trace other users.

On the whole, an adversary can try to perform the following attacks:

- Modify the offer description.
- Repudiate having issued a certain advertisement (when the adversary is the merchant).
- Remove the contribution made by some user to message dissemination.
- Issue a fake advertisement.
- Collect incentives corresponding to other users.
- Obtain the identity of a collaborating node and/or profile her by relating different interactions.

Attacks and security/privacy properties

Modification of an offer description. This attack refers to the integrity property. Offer descriptions are issued by the merchant, so we assume the merchant does not take part in the coalition. In our system, an advertisement consists of a message with the following structure:

$$\beta = \{\alpha,\ PubKeyChain,\ Multisignature,\ TimeChain\}$$

The advertisement itself is α which contains its public key certificate, the offer description and its expiration time:

$$\alpha = \{Cert_{Aut}\{PK_M\},\ Description,\ ExpirationTime\}$$

Integrity of the offer description is ensured since α is signed by the merchant (this signature is included in the $Multisignature$ field) and the signature scheme is unforgeable.

Advertisement repudiation. In our scheme, the merchant cannot repudiate having issued an advertisement since it has been signed and the signature on it is verifiable with a certified public key.

Note that, since collaboration in advertisement dissemination is anonymous, users do not need to repudiate having collaborated.

Removal of user contribution to dissemination. Another integrity aspect to be considered is whether users having contributed to the distribution of an advertisement can be unlawfully dropped and forgotten about. Let us assume an advertisement coming from merchant M that has been distributed by users $U_1, U_2, \ldots, U_n$. Let us assume that an intruder wishes to remove U_i from β. The intruder must remove the public key PK_{U_i} from $PubKeyChain$ and remove $\{Time \mid\mid \{Time\}_{SK_{U_i}}\}$ from $TimeChain$. Both removals can be done without any difficulty.

The difficulty for the intruder is to alter the $Multisignature$ field. This field contains the value

$$Multisignature = \mathcal{H}(\alpha)^{SK_M + SK_{U_1} + SK_{U_2} + \ldots + SK_{U_n}}.$$

The intruder must be able to obtain

$$Multisignature' = Multisignature \cdot (\mathcal{H}(\alpha)^{SK_{U_i}})^{-1}$$

Since discrete logarithms are hard to compute in a GDH group, the only way to obtain such value by an intruder is to get the $Multisignature$ field before U_i's contribution. This value can only be obtained if the intruder contacts directly the user who transferred β to U_i. This cannot be done due to the anonymity of the system.

Issuance of a fake advertisement. This attack refers to the authentication property. Our system requires the merchant to sign advertisements using a public key certified by an accepted authority. Generation of a certain advertisement that will be accepted as authentic coming from a valid merchant M requires knowledge of its private key SK_M. As long as this secret key is not compromised and the signature scheme is unforgeable (a valid signature can only be computed if the secret is known) the system provides authentication and remains secure against this attack.

Collecting incentives from other users. This situation refers to the authentication property too. In our system, E-coins given as incentives can only be collected by the users who have earned them. This is ensured by the incentive payment procedure. During this procedure, the merchant publishes the password required to obtain an e-coin encrypted with a public key whose corresponding private key is only known by the authentic user. In this way, only the authentic user will be able to obtain this password and request the e-coin.

Disclosure of the identity and/or tracing of users. This attack compromises the privacy of the users. This property consists of two components that must be guaranteed:

- *Anonymity*: Interaction with the system should not reveal the identity of the user.
- *Unlinkability*: It should not be possible to relate different interactions by the same user.

The anonymity of users collaborating in the dissemination of an advertisement is ensured because they simply are requested to provide a public key that does not reveal anything about their identity. Obtaining the password that permits to request an e-coin does not require the user to identify herself either. Finally, an anonymous e-coin system like [Chau89] also provides anonymity when obtaining and spending an e-coin.

Unlinkability is provided if users use a different key pair each time they perform an advertisement transfer. Each user U is able to randomly generate a new public/private key pair (SK_U/PK_U) at will and there is no connection between all the key pairs used by a certain user. Thus, two different public keys from the same user cannot be related by an observer.

Chapter 5

Private and trustworthy information spread in VANETs

Vehicular ad hoc networks allow vehicle-to-vehicle communication and, in particular, vehicle-generated announcements.

As explained in Section 1.1.3, *announcements* are spread to inform about road conditions (traffic jams, accidents). Vehicles which receive such announcements can take advantage of that information to select routes avoiding troublesome points. In contrast, *alert* messages are transmitted to warn nearby vehicles about dangerous movements (braking, lane change, etc). According to that, announcement messages require a longer dissemination range than alert messages. Besides, they demand a real-time processing which is much less strict than in the case of alerts. Therefore, advanced cryptography can be used to make such messages secure and trustworthy. Provided that the trustworthiness of such announcements can be guaranteed, they can greatly increase the safety of driving.

In this chapter, we present a new system designed to spread vehicle-generated announcements through VANETs. This work has been published

in [Daza08].

Trustworthiness is provided by following the *a priori* protection paradigm (see Section 2.3 for detailed discussion about protection paradigms). Internal attacks are thwarted by using an endorsement mechanism based on threshold signatures. Our system outperforms [Raya06b] in message length and computational cost. To the best of our knowledge, [Raya06b] is the most competitive scheme in the literature that follows the *a priori* protection paradigm.

Regarding privacy, we describe three different privacy-preserving variants of our system which ensure that vehicles volunteering to generate and/or endorse trustworthy announcements do not have to sacrifice their privacy (anonymity and unlinkability). The protocol detailed in [Raya06b] did not preserve the privacy of the volunteer vehicles.

Section 5.1 gives some cryptographic background needed to understand the proposed system. Section 5.2 presents our scheme in detail.

5.1 Cryptographic background

5.1.1 Secret sharing

A secret sharing scheme is a method by means of which a special figure, called dealer, distributes a secret s among a set $\mathcal{P} = \{P_1, \ldots, P_n\}$ of n players. The dealer secretly sends to each player P_i his share s_i of the secret s in such a way that only authorized subsets of players can recover the secret.

A (t, n)-threshold secret sharing scheme is a particular case in which authorized subsets are those composed of at least t players. Shamir's threshold secret sharing scheme [Sham79] gives a solution to this problem. Indeed, let $\mathbb{Z}_q$ be a finite field with $q > n$ and $s \in \mathbb{Z}_q$ be the secret to be shared. The dealer picks a polynomial $p(x)$ of degree at most $t-1$ at random, whose free

term is the secret s, that is, $p(0) = s$. The polynomial $p(x)$ can be written as $p(x) = s + \sum_{j=1}^{t-1} a_j x^j$, where $a_j \in \mathbb{Z}_q$ has been randomly chosen.

Each player P_i is assigned a known value $\alpha_i \in \mathbb{Z}_q$. Then, the dealer privately sends to player P_i his share $s_i = p(\alpha_i)$, for $i = 1, \ldots, n$.

Therefore, a set $A \subset \mathcal{P}$ of at least t players can recover the secret $s = p(0)$ by interpolating the set of shares they hold:

$$s = p(0) = \sum_{P_i \in A} s_i \lambda_i^A = \sum_{P_i \in A} s_i \left(\prod_{P_j \in (A \setminus P_i)} \frac{-\alpha_j}{\alpha_i - \alpha_j} \right)$$

Values λ_i^A are called the Lagrange coefficients. It can be proven that less than t players cannot obtain any information about the secret s.

5.1.2 Threshold signatures

Digital signatures allow to send authenticated and non-repudiable messages. The message sender is required to have a public/private key pair. Signature generation is an algorithm that takes as input the message, m, and the sender's private key, SK. Its output is the signature $\sigma(m)$ on m. Signature verification is performed by the receiver. Its algorithm takes as input the message m, its signature $\sigma(m)$ and the sender's public key PK. It outputs "yes" or "no" to reflect the validity of $\sigma(m)$. A valid signature convinces the receiver about the integrity of m and is taken as a proof that the message was generated by the authentic sender (the only party knowing SK).

A (t, n)-threshold signature distributes the signing operation among a group of n participants. Each participant in a distributed signature scheme is given a share, SK_i, of the secret key, SK, in such a way that to sign a message every participant computes a partial signature, $\sigma_i(m)$, using his share of the secret key. Then, any set of at least t participants can compute a valid

signature $\sigma(m)$ on the message by combining their partial signatures. The resulting signature is equivalent to the one that results in the non-distributed case (it is also verifiable using PK). A distributed signature scheme is said to be non-interactive if every participant can compute his partial signature on a message m without interacting with the rest of participants. Signatures in [Shou00, Bold03, Fouq01, Damg01] are examples of non-interactive threshold signature schemes.

For the sake of concreteness, we next recall an efficient threshold signature scheme, namely the one in [Bold03], a distributed version of the signature scheme by Boneh, Lynn and Shacham (BLS, [Bone01]). Both schemes work over Gap Diffie-Hellman (GDH) groups – see original papers for more details. In a nutshell, these signature protocols based on pairings are quite efficient as the signing process only requires hash operations and modular exponentiations and the verification process two pairing computations. In [Barr02] a fast implementation of the Tate pairing computation was given and the BLS signature scheme was compared with an RSA signature on a Pentium PIII processor at 1 GHz. Using RSA with a modulus length $|n| = 1024$ bits and a private exponent length $|d| = 1007$ bits, signing took 7.90 ms and verifying took 0.4 ms. Using the BLS signature with elliptic curves over $\mathbb{F}_{3^{97}}$, signatures were 160 bits long (which yields a similar security as the above-mentioned 1024-bit RSA signature), and signing and verifying took 3.57 ms and 53 ms, respectively. So there exist threshold signatures with reasonable computational cost.

Let $\mathbb{G}$ be a GDH group, $g =< \mathbb{G} >$ be a generator of the group and p be the order of the group. Using methods described in [Genn96], every participant P_i obtains a share SK_i. The set of shares realizes a (t, n)-threshold access structure, that is, t parties can retrieve the secret key SK whereas

less than t cannot obtain any information on the secret key. The retrieval process can be performed by means of Lagrange interpolation and also yields the matching public key $PK = g^{SK}$. To sign a message m, a participant P_i computes his partial signature as $\sigma_i(m) = \mathcal{H}(m)^{SK_i}$ ($\mathcal{H}$ is a public one-way and collision-free hash function) and broadcasts $\sigma_i(m)$. After a set A of at least t participants have broadcast their partial signatures $\sigma_i(m)$ for message m, a standard signature σ for the message can be computed as

$$\sigma(m) = \prod_{i \in A} \sigma_i(m)^{\lambda_i^A} = \mathcal{H}(m)^{\sum_{i \in A} \lambda_i^A SK_i} = \mathcal{H}(m)^{SK}$$

where λ_i^A are the Lagrange coefficients.

5.1.3 Privacy in secret sharing

In short, an anonymous secret sharing scheme is one where participants can co-operate in the retrieval of the secret while keeping their identity undisclosed (anonymity) and without successive co-operations by the same participant being linkable (unlinkability). Shamir's (t, n)-threshold secret sharing scheme described in Section 5.1.1 does not offer unlinkability: each Lagrange coefficient corresponds to a certain participant P_i and, even if that correspondence is kept secret for anonymity (*i.e.* by using the underlying α_i as pseudonyms), successive co-operations by the same participant can be linked because the Lagrange coefficient of the participant appears every time. Anonymous secret sharing schemes in the literature present a very high cost that limits their practical applicability [Blun97].

Note that, if the secret sharing scheme underpinning a threshold signature protocol is not anonymous, the resulting threshold signature is either linkable (successive partial signatures by a participant can be linked) or requires a

trusted third-party and is thus unsuitable for a VANET.

5.2 Trustworthy privacy-preserving announcements in VANETs

In this section, a new system for secure announcements in VANETs is presented. It uses digital signatures to prevent external attackers from being able to inject false messages and follows the *a priori* approach to thwart fake announcements sent by internal attackers. An announcement will only be considered as being valid if it has been endorsed by at least t different vehicles.

5.2.1 Non-private protocol

For clarity, let us begin with a protocol which can offer anonymity but not unlinkability.

- *Set-up:* During this stage, the carmakers set up a (t, n)-threshold signature scheme, where n is the maximum number of vehicles allowable in the VANET. To do this, the carmakers must agree on a polynomial of degree $(t-1)$ that will be evaluated at points α_i, for $i = 1$ to n. The range of n points is partitioned into several subranges, each of which is assigned to a carmaker. The number n can be very large without scalability problems. Next, a public key PK and n shares $SK_i, i = 1, \ldots, n$ of the secret key SK are generated. Each vehicle P_i is equipped with the public key PK and its secret key share SK_i; the share SK_i is held in a smart card plugged into the vehicle (tamper-resistance is assumed for the card in what follows). When input the hash value $\mathcal{H}(m)$ of a

message, the smart card returns a partial signature on that hash value, that is, $\sigma_i(m) = \mathcal{H}(m)^{SK_i}$. Anonymity is obtained by not linking SK_i with the identity of the vehicle; this makes sense for other reasons too because, smart cards being removable, several smart cards each holding a different secret key share could alternatively be used with the same vehicle (like several cards can be used with a cellphone).

- *Announcement generation:* When a vehicle P_i wishes to send an announcement m, P_i computes the partial signature $\sigma_i(m)$ and broadcasts m and $\sigma_i(m)$. An announcement should only reach vehicles that are close enough to the originating vehicle so as to be able to check the validity of the announced condition. Since they do not need to reach distant points, announcement messages are not relayed by VANET nodes and they travel only up to the range of the broadcast technology used (even if a maximum range of 1000 meters for car-to-car communication with the Dedicated Short Range Communication protocol is reported in [Raba07], typical ranges from 300 to 500 meters on highways and about 100 meters in cities are mentioned in [Berg07]).

- *Announcement endorsement:* If vehicle P_j receives an announcement m (together with the partial signature on it by the announcement originator P_i) and wishes to endorse m, then P_j computes its own partial signature $\sigma_j(m)$ on m and broadcasts $\mathcal{H}(m)$ and $\sigma_j(m)$ to return them to P_i, where $\mathcal{H}()$ is the same hash function used in the signature computation. As in announcement generation, messages with partial signatures are not relayed.

- *Signature composition:* The vehicle P_i which generated an announcement stores m and the partial signatures on m it receives (partial signatures on m are identifiable by the hash $\mathcal{H}(m)$ they carry). Once P_i has collected t different partial signatures on m, it can compute a standard signature $\sigma(m)$ and broadcast it along with m.

- *Announcement reception and verification:* Vehicles in the VANET will only consider as trustworthy those announcements carrying a standard signature that can be verified using the public key PK. The use of the threshold signature scheme provides vehicles with the assurance that such a standard signature can only have been computed if at least t vehicles have endorsed m by computing their partial signature on it. These messages, containing a standard signature, will be relayed by VANET nodes. In this way, they will reach distant vehicles which will benefit from the information in the messages.

The reason for keeping SK_i in a smart card is to prevent the vehicle driver from learning SK_i; otherwise, t colluding drivers could recover the secret key SK, which would allow any single one of them to sign messages that would be accepted as trustworthy without any endorsement.

In any case, the choice of t is a trade-off between security and availability. On one hand, t should be high enough so that the probability of there being t or more within-range colluding vehicles who could validly endorse fake messages is reasonably low (security). On the other hand, t should not be so high that finding $t-1$ additional within-range endorsers is too difficult for an honest announcement generator (availability).

The problem with the above protocol is that it lacks privacy and, more precisely, unlinkability. This is due to the fact that signature composition requires computing Lagrange coefficients (see Section 5.1.1). Computation

of such coefficients requires in turn knowledge of the value α_i assigned to each vehicle P_i having contributed a partial signature. Certainly, it can be assumed and it is assumed that the correspondence between P_i and α_i is withheld (α_i is used a pseudonym for vehicle P_i), which provides anonymity. However, different partial signatures generated by the same vehicle P_i all use α_i, so they are linkable. Therefore, unlinkability is not achieved.

5.2.2 Cost analysis of the non-private protocol

In this section we compare the cost of our non-private protocol above with the cost of the concatenated signatures protocol in [Raya06b]. Both protocols are non-private, so the comparison is fair. In the next subsections, the cost is analyzed in terms of announcement length, announcement generation time and announcement verification time.

Announcement length

In the concatenated signatures protocol in [Raya06b], authenticated announcements contain as many signatures and public key certificates as endorsing vehicles, so their length is $O(t)$. In our proposal, both the partially signed announcements and the completely signed announcements contain a single signature, so the length of announcements is $O(1)$.

Since the above comparison in O-notation may be misleading for small values of t, we next compare both proposals by taking the constant terms into account. We assume that [Raya06b] uses the concatenated signatures protocol with the RSA public key cryptosystem with 1024 bit moduli (so, digital signatures will be 1024 bits long). Let us consider that the information which is announced is a bits long. The concatenated signatures protocol in [Raya06b] requires one signature and one public key certificate from t

different signers. We will consider that a digital certificate contains an RSA public key (barely longer than the 1024-bit modulus if a short public exponent is used), the owner's pseudonym (which could be a 64-bit serial number) and a signature by the Certification Authority (1024 bits). According to that, the total length of an announcement in [Raya06b] is $a+t\cdot(3\cdot 1024+64)$ bits. For example, if four endorsing vehicles are required ($t = 4$), this scheme yields an announcement length of $a + 12544$ bits. With the same assumptions, our proposal has a constant announcement length of $a + 160$ bits (we are using the BLS signature scheme). As t grows, the advantage of using our system increases.

Announcement generation delay

In [Raya06b] vehicles sequentially contribute with their signature to endorse an announcement. This means that a valid message generation takes at least the time necessary for a message to perform $t-1$ hops plus the time required to compute t digital signatures. This is an $O(t)$ cost. Let j be the time (in milliseconds) necessary for a message to perform one hop. According to the signature generation time reported in Section 5.1.2, a valid message generation in [Raya06b] using the RSA cryptosystem with 1024 bit public keys takes $7.90 \cdot t + (t-1) \cdot j$ ms.

In our protocol, vehicles can endorse a message in parallel. So, the delay due to data transmission required to generate a valid message is fixed to the time to perform 2 hops (one from the generator to within-range endorsers and another from endorsers to the generator) plus the time to compute 2 BLS signatures. This time is $2 \cdot (j + 3.57)$ ms.

After t endorsement messages have been collected in our protocol, a standard signature is composed by the vehicle originating a message at an $O(t)$

cost (the cost of computing a standard signature from t partial signatures). As can be seen in Section 5.1.2, the cost of this operation is dominated by the exponentiation of each partial signature to its corresponding Lagrange coefficient. The cost of each exponentiation is similar to the cost of computing one digital signature (also consisting of one exponentiation). Thus, the composition time is approximately $t \cdot 3.57$ ms.

The overall generation time with our protocol is $2 \cdot (j + 3.57) + t \cdot 3.57$ ms. This expression can be rewritten as $2 \cdot j + (t + 2) \cdot 3.57$ ms. This is a shorter time than the one required by [Raya06b]. As t grows, the advantage of using our system increases.

Announcement verification time

In [Raya06b] announcement verification requires checking t signatures and t public key certificates. If certificates are subject to revocation, there is an additional cost related to checking certificate revocation lists (even this cost is not explicitly mentioned in [Raya06b]). In any case, the verification cost is $O(t)$.

In our protocol, an announcement is verified by checking one signature. Since the public key PK used for verification is always the same and is stored in the smart card by the carmaker, its validity does not need to be checked. This is an $O(1)$ cost.

Let us now consider the constant terms for greater accuracy. Assume the RSA and the BLS signature schemes are used by [Raya06b] and our proposal, respectively. Section 5.1.2 details the signature verification time for each signature scheme. In this way, an announcement verification in [Raya06b] takes $2 \cdot 0.4 \cdot t$ ms (the verifier checks the certificate and message signatures sent by each endorsing vehicle). The same operation using our protocol takes 53

ms. Therefore, with those assumptions, our proposal outperforms [Raya06b] only when $t \geq 67$. In practice, t will be usually less than 67, so that [Raya06b] will normally be faster than our protocol as far as the computation involved in signature verification goes.

Nonetheless, if the time and communication needed to check certificate revocation lists was taken into account, our proposal would be more efficient, because in [Raya06b] a certificate revocation list may need to be checked for each certificate to be verified.

Summary of cost analysis

Table 5.1 summarizes the cost of both protocols as a function of the threshold t. The strong points of our proposal are that the following is constant: announcement length and announcement verification time.

If a more accurate analysis of the constant terms is performed (which is necessary when t is small), it turns out that our system still yields shorter announcements and faster announcement generation than [Raya06b]. Announcement verification, on the contrary, is faster with [Raya06b] at least for the usual (small) values of t.

However, if the cost of checking certificate revocation lists is considered in announcement verification, the picture changes dramatically. Indeed, [Raya06b] requires verifying t certificates, which may require checking certificate revocation lists t times. This may be very long, as it involves communication, not just computation. In our proposal, the validity of PK does not need checking, as explained above. So, when the cost of checking certificate revocation is included, our proposal is more efficient also for announcement verification.

Table 5.1: Cost breakdown as a function of the threshold t of the non-private protocol in [Raya06b] and the non-private protocol in this paper

	Protocol [Raya06b]	Our protocol
Announcement length	$O(t)$	$O(1)$
Announcement generation time	$O(t)$	$O(t)$
Announcement verification time	$O(t)$	$O(1)$

5.2.3 Group-based private protocol

In this section, a modification of the previous protocol is described in order to provide unlinkability. The modification mainly affects the set-up phase.

- *Set-up:* The n vehicles that form the VANET are divided into r groups, with each group consisting of n/r vehicles (for simplicity, it is assumed that parameters n and r are chosen so that r divides n, but suitable rounding can be used in the general case). The carmakers set up a (t, r)-threshold signature scheme. During this generation, a public key, PK, and r shares, $SK_j, j = 1, \ldots, r$, of the secret key SK are generated (one share for each group). Each carmaker keeps a copy of each of the r shares. Each manufactured vehicle P_i is randomly assigned by the carmaker to a group j; then it is equipped with the public key PK and the secret key share SK_j assigned to its group (as above, SK_j is held in a smart card plugged to the vehicle).

This modification causes vehicles belonging to the same group to be assigned the same secret key share. In this way, partial signatures cannot be related to a single vehicle but to any member of its group. If groups are large enough, this protocol provides unlinkability. On the other side, a valid signature $\sigma(m)$ must now be generated not just by any t vehicles, but by vehicles belonging to at least t *different* groups.

Security, privacy and availability

Parameters t and r of the group-based protocol have an impact on security against fake messages, on privacy and on availability.

The threshold t should be set high enough so that the probability of there being t or more colluding vehicles who could validly endorse false announcements is reasonably low.

For a choice of t, parameter r must be chosen considering the trade-off between unlinkability and availability:

- *Unlinkability.* The group size $g := n/r$ must be large enough so that linkability at the group level (which cannot be avoided) does not imply linkability at the vehicle level.

- *Availability.* The number of groups r must be large enough so that, given an announcement, finding t endorsing vehicles from different groups is easy. Thus, $r \gg t$.

By construction, this proposal has the same cost as the non-private protocol (see Section 5.2.1).

5.2.4 Extended group-based private protocol

In the previous group-based protocol, it may be difficult in some cases to find a value for r striking a balance between unlinkability and availability. This is the case when the VANET is sparse or consists of an actual number n' of vehicles much less than the maximum allowable number n. Since the group size cannot be too small if unlinkability is to be preserved, the number r of groups has to be small. In those conditions finding t within-range endorsing vehicles from different groups may be quite challenging.

A solution to mitigate the problem caused by a small r is to use d different threshold signature schemes so that, if t within-range endorsing vehicles from different groups cannot be found for the first scheme, they are sought for the second scheme, and so on. The modified set-up, announcement generation, endorsement and signature composition phases are:

- *Set-up:* The n vehicles that form the VANET are divided into r groups, as in Section 5.2.3. The carmakers set up d different (t, r)-threshold signature schemes. For $k = 1$ to d, the k-th scheme consists of a public key PK^k and r shares, SK^k_j, $j = 1, \ldots, r$ (one share per group). Each carmaker keeps a copy of all r shares for all d signature schemes. For $i = 1, \ldots, n$, each manufactured vehicle P_i is equipped with the public keys $(PK^1, \ldots, PK^d)$ and the secret key shares $(SK^1_{i_1}, \ldots, SK^d_{i_d})$, where $i_k \in_R \{1, \ldots, r\}$ is the group randomly assigned by the carmaker to P_i for the k-th threshold signature scheme. As above, all secret key shares are held in a smart card.

The only variation in the extended group-based protocol with respect to the previous protocols (non-private, group-based) in what respects the announcement generation, endorsement and signature composition steps is that now messages in those steps must include a field specifying which threshold signature scheme among the d possible ones is being used in a particular execution.

Announcement generation, endorsement and signature composition are attempted for the first threshold signature scheme as in Section 5.2.3. If, after a predefined timeout, partial signatures from t different groups have not been collected, announcement generation and endorsement are re-started for the second threshold signature scheme. The process stops when a threshold

signature scheme is found for which endorsements from t different groups can be collected. In the worst case, all d threshold signatures schemes can fail.

Storage requirements at the vehicles are increased. In this case, each vehicle stores d key shares and d public keys (compared to one share and one public key in the previous proposal).

5.2.5 Semi-private protocol for sparse VANETs

The protocol in Section 5.2.4 is not without drawbacks. Even with d different threshold signature schemes, collecting endorsement from t different groups may fail in very sparse VANETs. A way to circumvent the above problem is to drop groups but to keep several threshold signature schemes for privacy. The modified protocol looks as follows:

- *Set-up:* The carmakers set up d' different (t, n)-threshold signature schemes. Like in the non-private protocol of Section 5.2.1 but for each signature scheme in this protocol, the range of n points corresponding to possible vehicles is partitioned into several subranges, each of which is assigned to a carmaker. For $k = 1$ to d', the k-th scheme consists of a public key PK^k and n shares, SK_i^k, $i = 1, \ldots, n$ (one share per vehicle). For $i = 1, \ldots, n$, each vehicle P_i is equipped with the public keys $(PK^1, \ldots, PK^{d'})$ and the secret key shares $(SK_i^1, \ldots, SK_i^{d'})$, with share SK_i^k being obtained by evaluating the polynomial of the k-th scheme at point α_i^k, where α_i^k is assumed to belong to the subrange of the carmaker of P_i for the k-th scheme.

- *Announcement generation:* When a vehicle P_i wishes to send an announcement m, P_i randomly selects one of the d' threshold signature schemes, say scheme k. One can assume that the selection is performed

by the smart card in the vehicle so that the selected k is beyond the user's control. Then P_i computes its partial signature $\sigma_i^k(m)$ on m and broadcasts the announcement and its partial signature. This solution also requires messages to include a field indicating which signature scheme k is being used.

- *Announcement endorsement:* If vehicle P_j receives the announcement m (together with the partial signature on it by the announcement originator P_i) and wishes to endorse m, P_j uses the k-th threshold scheme to compute its own partial signature $\sigma_j^k(m)$ on m and broadcasts $\mathcal{H}(m)$ and $\sigma_j^k(m)$, where $\mathcal{H}()$ is the same hash function used in the signature computation.

- *Signature composition:* The vehicle P_i which generated an announcement stores m and the partial signatures on m it receives (partial signatures on m are identifiable by the hash $\mathcal{H}(m)$ they carry). Once P_i has collected t different partial signatures on m, it can compute a standard signature $\sigma^k(m)$ and broadcast it along with m.

- *Announcement reception:* Vehicles in the VANET will only consider as trustworthy those announcements carrying a standard signature that can be verified using a public key PK^k in the set $(PK^1, \ldots, PK^{d'})$.

The above semi-private protocol requires vehicles to store d' shares and d' public keys.

Security, privacy and availability

As in the previous protocols, the threshold t is the parameter controlling security against insertion of fake announcements.

Unlinkability is related to parameter d', the number of threshold signature schemes set up by the carmaker for this protocol. Provided that the threshold signature scheme is randomly selected, the probability that two successive participations by P_i can be linked is $1/d'$ (this happens if the same threshold signature scheme is selected in both cases). Thus, unlinkability improves with respect to the non-private protocol (Section 5.2.1) but it is worse than in the group-based or extended group-based protocols (Section 5.2.3 and 5.2.4, respectively). However, the advantage is increased availability in that there are no constraints on the t vehicles that must endorse an announcement (any t vehicles will do), so that the endorsement process is easier in very sparse VANETs with really few vehicles per area unit.

A way to improve unlinkability is by taking a large d', which does not affect the announcement verification time. This is different from what happens in the extended group-based protocol if parameter d is increased: there, the signature schemes are tried one after the other until a valid signature is obtained or the d schemes have been tried, so a large d may result in longer verification times.

5.2.6 Compound protocol

The protocol in Section 5.2.5 can be used as a fallback for the protocol in Section 5.2.4, which in turn is a fallback for the protocol in Section 5.2.3. The idea is that vehicles can be set up for all three protocols by the carmaker. The first option to be tried is the group-based protocol. If traffic sparseness is such that partial signatures from t different groups cannot be collected for a certain announcement before a fixed timeout, then the protocol in Section 5.2.4 is used. If this does not work either, the protocol Section 5.2.5 can be used to get limited unlinkability without increasing the difficulty of

collecting endorsements with respect to the non-private protocol. According to that, a compound protocol combining the protocols in Sections 5.2.3, 5.2.4 and 5.2.5 can be specified as follows:

1. Initially, the group-based protocol of Section 5.2.3 is used. Note that this protocol is a particular case of the extended group-based protocol where there is only one (t, n)-threshold signature scheme in use. Thus, hereafter we will consider this step as a part of the next step, where the extended group-based protocol is used. (In what follows we will only refer to the extended group-based protocol and the semi-private protocol. This also applies to the simulation results which will be presented in Section 5.2.7.)

2. If a complete signature cannot be constructed before a certain timeout, the extended group-based system of Section 5.2.4 is launched. Construction of a complete signature by means of d different (t, n)-threshold signature schemes is attempted. The timeout in use depends on the value t (the number of different partial signatures required to compute a standard signature). For each unit increase of threshold t, the timeout increases by β milliseconds.

   Each of the d signature schemes is tried in sequence until a complete signature is constructed or the timeout expires, so at most $d \times timeout$ milliseconds are spent on the extended group-based system.

3. If the extended group-based system does not work either, the semi-private protocol (see Section 5.2.5) is tried.

Compound set-up phase

The compound protocol is composed of three schemes. In previous sections, we have presented the set-up phase for each of these schemes. We next explain the compound set-up phase when deploying such a system in a real environment.

Let us consider the co-existence of m carmakers in a certain area. The i-th carmaker produces v_i hundreds of thousands of vehicles per year. According to the European Environment Agency [Eea08], the EU-15 area had about 170 millions of vehicles in 2004. Even though the carmakers produce $v_1 + \cdots + v_m$ hundreds of thousands of new cars each year, there is also a large quantity of old vehicles which are eliminated in the same period. Therefore, the size of the vehicle fleet in a certain area does not undergo a strong increase from year to year. Value n is the maximum number of vehicles allowable in the system covering the area. The only assumption on n is that it cannot be greater than the cardinality of the group used to construct the BLS signature scheme. For cryptographic security reasons, this cardinality should be at least 2^{160}. So, we can set a value for n close to this upper limit. Such a huge n ensures that we will never run out of key shares. As it can be seen in Section 5.1 a huge n can be used without any negative impact on the system performance.

Our system requires a governmental authority GA in the geographical area of deployment to ensure a correct set-up phase. Note that this authority is no longer needed when executing the compound protocol. The only role of the authority is to coordinate share distribution among the vehicles produced by different carmakers. In this way, GA establishes d signature schemes and the number r of groups of vehicles in the area. According to that, each signature scheme generates r shares. Each share is linked to one

group. Additionally, GA partitions the n possible vehicles into several subranges, each of which is assigned to a carmaker. It also establishes d' different threshold signature schemes. Each one generates n shares. Note that a certain carmaker receives the shares that correspond to its assigned subrange of n.

Now, let us consider that a certain vehicle P_i is manufactured. This car has to be set up by its carmaker for both extended group-base and semi-private protocols. This process has been explained individually in Sections 5.2.4 and 5.2.5. We next summarize it:

- *Extended group-base protocol.* For each signature scheme $k = 1, \ldots, d$, P_i is randomly assigned by the carmaker to a group j_k (where $j_k \in \{1, \ldots, r\}$) and it is equipped with the share corresponding to group j_k.

- *Semi-private protocol.* P_i is equipped with shares $(SK_i^1, \ldots, SK_i^{d'})$, where the share SK_i^w is obtained by evaluating the polynomial of the w-th scheme at point α_i^w, which is assumed to belong to the subrange of the carmaker of P_i.

The compound set-up phase we have presented relies on the assumption that an authority GA exists which coordinates share distribution. An open problem is to devise a compound set-up phase which can work when no GA is available.

5.2.7 Simulation

Our scheme for secure vehicle-generated announcements over VANETs was simulated in a realistic environment, where the range of car-to-car broadcasts was assumed to be 100m (the worst-case, urban range according to [Berg07]).

The goal of our simulations is to observe the performance of the compound protocol explained in Section 5.2.6. This protocol requires a timeout that depends on values β and t. In our simulations, we have fixed β to 50 ms. According to that, $t = 4$ represents a timeout of 200 milliseconds.

Simulation set-up

The network simulator *ns-2* [Netw08] was used. The VANET scenario was built using the scenario generator presented in [Saha04]. The road network considered covered an area of 2.4 km by 2.4 km and is shown in Figure 5.1.

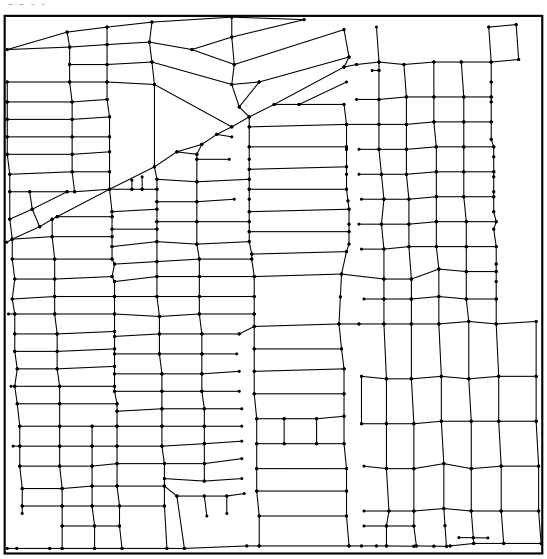


Figure 5.1: Simulation scenario

In our simulations, the primary indicator examined is the probability for a certain announcement to get validated. An announcement is validated when its standard signature is constructed from t different partial signatures generated by t different cars. Those vehicles can belong to t different groups or to only one group depending on whether the extended group-based protocol or the semi-private protocol are used.

A second indicator taken into account is the average number of different (t, n)-threshold signature schemes which are used when applying the extended group-based protocol. This indicator determines the time needed to validate the announcements.

Both indicators are essential to evaluate whether our scheme is usable in real VANETs.

In the next subsection, the optimal values for the parameters used in our system are studied. The goal of that study is to select the values for parameters based on the two indicators stated above for a wide range of vehicle densities. When the system is running, parameter values cannot be easily modified, so a parameter choice must be made which works well under several road conditions. The lessons learned from the simulations are summarized in the last subsection.

The results given in what follows are average values obtained from 100 executions performed for each parameter choice.

Parameter selection

Vehicle density is expressed in $vehicles/km^2$. This value is changed by varying the total number of vehicles in the scenario represented in Figure 5.1.

Let t stand for the minimum number of vehicles needed to validate an announcement. Each vehicle should belong to a different group when using the extended group-based protocol. Under the semi-private protocol there are no group constraints.

Table 5.2 shows the average probability p_1 of a certain announcement to be validated using the extended group-based protocol for fixed r and several

Table 5.2: Average validation probability p_1 and average number i of different threshold signature schemes tried for the extended group-based protocol, for constant number of groups $r = 10$. Average validation probability p_2 for the semi-private protocol when the extended group-based protocol fails. Results are given as a function of vehicle density and the minimum number of validating vehicles t.

Vehic. dens.	$t = 4$			$t = 5$			$t = 6$		
	p_1	i	p_2	p_1	i	p_2	p_1	i	p_2
6.94	0.48	1.67	0.07	0.24	1.81	0.12	0.00	N/A	0.04
8.68	0.64	1.69	0.02	0.33	1.74	0.05	0.04	2.00	0.04
12.15	0.70	1.25	0.01	0.40	1.80	0.04	0.20	1.60	0.11
15.62	0.72	1.22	0.00	0.44	1.70	0.04	0.36	1.85	0.10
17.36	0.76	1.17	0.00	0.68	1.59	0.09	0.52	1.57	0.08
24.31	0.94	1.14	0.00	0.76	1.53	0.02	0.72	1.67	0.11
31.25	0.96	1.13	0.00	0.92	1.17	0.00	0.81	1.51	0.05
38.19	0.96	1.09	0.00	0.94	1.08	0.00	0.89	1.48	0.00
45.14	1.00	1.00	N/A	0.94	1.09	0.00	0.92	1.33	0.00
52.08	1.00	1.00	N/A	1.00	1.00	N/A	0.96	1.26	0.00

values of t. Value i indicates the average number of different threshold signature schemes tried (each one is used until a timeout occurs) in order to validate the announcement with the extended group-based protocol. When the number of different threshold signature schemes tried for a certain announcement reaches the number d of available schemes without the announcement being validated under any of them, the semi-private protocol is launched. We have set $d = 3$ given the values i obtained in preliminary simulations. This will be further explained below.

Value p_2 is the average probability of validating the announcement under the semi-private protocol when the extended group-based protocol fails. Note that the semi-private protocol is a tolerable fallback for low vehicle densities. For higher vehicle densities, the probability p_1 of successful validation with the extended group-based protocol is already very high, so that the instances in which the semi-private protocol is used as a fallback are very difficult ones (*e.g.* very sparse locations); this explains the near zero p_2 values for higher densities. Also, the N/A value for p_2 means that there was no need to call the semi-private protocol. We have set $d' = 20$ as the number of different threshold signature schemes available in the semi-private protocol; this should yield a good trade-off between unlinkability and implementation cost in the vehicles.

Results in Table 5.2 are given as a function of vehicle density and the minimum number of validating groups t. For this experiment, the number of groups of vehicles was set to $r = 10$. The dependency on this value r will be studied below, in Table 5.3.

It can be observed in Table 5.2 that both validation probabilities p_1 and p_2 decrease as t increases, no matter whether the VANET is sparse or dense. This is not surprising because validation is "easier" for smaller t; however,

the price paid is that for smaller t the trustworthiness of a validated message is lower. Following this argument, it is also expected that for very sparse networks (vehicle density of 6.94) and high t values ($t = 6$ for instance) the extended group-based protocol is unable to validate a single announcement; in fact, not even the semi-private protocol works properly in that setting ($p_2 = 0.04$ for a vehicle density of 6.94). As a trade-off between trustworthiness and availability, it is suggested to take $t = 4$ or $t = 5$ depending on the desired trustworthiness level for the announcements. In fact, $t = 5$ is the highest reasonable value because, even though $t = 6$ works fine for dense VANETs (vehicle density above 38.19), it does not for medium-density ($p_1 = 0.52$ for a density of 17.36) and sparse VANETs. Since a threshold must be chosen which works properly under several road conditions, it is better to select $t < 6$. In what follows, $t = 4$ is taken.

Simulation shows that the average number i of different threshold signature schemes tried by the extended group-based protocol decreases when the vehicle density increases and increases when the threshold t increases. All in all, usually $i \leq 2$ whenever validation is successful, which is the usual outcome for medium- to high-density VANETs and moderate threshold ($t = 4$). For very sparse networks, validation mainly relies on the semi-private protocol so we can choose the number d of signature schemes for the extended group-based protocol by considering only medium- to high-density VANETs. Thus a choice of $d = 3$ is fair enough and is assumed in what follows; this implies that at most $3 \times timeout$ milliseconds are spent on the extended group-based protocol (as said above, for $t = 4$ we consider $timeout = 200$ milliseconds, so the overall time spent on the extended group-based protocol is 600 ms).

Table 5.3 shows the average probability p_1 of a certain announcement

Table 5.3: Average validation probability p_1 for the extended group-based protocol and average validation probability p_2 for the semi-private protocol when the extended group-based protocol fails; average group size g is shown too. Results are given as a function of vehicle density and number of groups r, for constant threshold $t = 4$

Vehic. dens.	$r = 8$			$r = 10$			$r = 15$		
	p_1	p_2	g	p_1	p_2	g	p_1	p_2	g
6.94	0.33	0.08	5.0	0.48	0.07	4.0	0.54	0.01	2.7
8.68	0.38	0.05	6.2	0.64	0.02	5.0	0.66	0.03	3.3
12.15	0.52	0.04	8.7	0.70	0.01	7.0	0.76	0.00	4.7
15.62	0.67	0.00	11.2	0.72	0.00	9.0	0.76	0.00	6.0
17.36	0.71	0.00	12.5	0.76	0.00	10.0	0.88	0.00	6.7
24.31	0.86	0.00	17.5	0.94	0.00	14.0	0.96	0.00	9.3
31.25	0.93	0.00	22.5	0.96	0.00	18.0	1.00	N/A	12.0
38.19	0.94	0.00	27.5	0.96	0.00	22.0	1.00	N/A	14.7
45.14	0.97	0.00	32.5	1.00	N/A	26.0	1.00	N/A	17.3
52.08	1.00	N/A	37.5	1.00	N/A	30.0	1.00	N/A	20.0

being validated using the extended group-based protocol for fixed t and several values of r. If this protocol fails, the semi-private fallback is launched. Value p_2 represents the average probability of validating a message with the semi-private protocol when the extended group-based protocol fails. Finally, value g represents the average group size. All results are given as a function of vehicle density and the number of groups of vehicles r. A value of r must be set which works fine for very different vehicle densities. Also, r must be chosen considering the trade-off between unlinkability and availability (see related discussion in Section 5.2.3): value r should be greater than t in order to guarantee availability (*i.e.*, so that finding t endorsing vehicles from different groups is easy). However, a big r implies that the group size g is small ($g := n/r$). In this way, the unlinkability of a certain vehicle is poor. In contrast, for a small r, the unlinkability of a certain vehicle is very high but the

validation probability decreases. Table 5.3 reflects the availability problems of the system in very sparse VANETs when a certain unlinkability level is demanded. More specifically, we can observe that with a vehicle density of 6.94 and an average group size $g = 5.0$ (which occurs when $r = 8$), the probability p_1 of a certain announcement to be validated is 0.33. Note that larger group sizes (which imply $r \ll 8$) will yield worse availability results. When availability problems arise, the system resorts to the semi-private protocol (which is less good in terms of privacy, unless d' is extremely high).

According to the above considerations, $r = 10$ is taken as a reasonable trade-off between unlinkability and availability for all vehicle densities.

Note. As mentioned in Section 5.2.3, unlinkability is proportional to the group size g. One might object that the average group size in the simulations is small, which is true because the small geographical area considered (2.4 km by 2.4 km) can only accommodate a small number of vehicles. However, the purpose of the simulation is to evaluate the validation probability, which is independent of the group size (it only depends on the threshold t, the number of groups r and the vehicle density). In a real scenario (*e.g.* the EU-15 area with 170 million vehicles mentioned in [Eea08], the same t and r values used in the simulations can be employed, which will result in a very large group size g guaranteeing high unlinkability.

Lessons learned from the simulation

The probability of successful validation depends on the threshold t, the number r of groups and the vehicle density, regardless of the group size. For a fixed density, the greater r with respect to t, the higher the success probability. The closer r to t, the lower the success probability.

All simulations performed reflect that with the parameter selection used

($t = 4$, $r = 10$ and $d = 3$), our proposal provides message trustworthiness and vehicle unlinkability under different road conditions. Results show that our scheme performs best in medium- to high-density VANETs (densities from 12.15 to 52.08). Nevertheless, it works fair enough in very sparse environments as well:

- For a vehicle density 6.94, our scheme achieves a success probability $p_1 = 0.48$ in announcement validation with the extended group-based protocol. In the cases when this protocol fails, the semi-private one works with a probability $p_2 = 0.07$.

- For a vehicle density 8.68, the success probability with the extended group-based protocol increases to $p_1 = 0.64$. The semi-private protocol used as a fallback earns an additional $p_2 = 0.02$.

The low success in validation for sparse VANETs should be put in context: in an area with very low traffic, it is often less critical to get announcements on road conditions, as there is hardly anyone who can benefit from them.

Chapter 6

Private resource access in social networks

In Section 1.1.4, we pointed out the need to design privacy-preserving resource access protocols for social networks.

Regarding this topic, the latest proposals [Carm07, Domi07] in the literature provide private resource access in social networks by enabling private relationships between the users of the network. Even though [Domi07] overcomes the limitations detected in [Carm07], it also has some shortcomings that should be solved, which were discussed in detail in Section 2.4. We next summarize them:

- In that system, a certain user with a small number of relationships is likely to stay isolated at certain periods of time (e.g. early in the morning). At these periods, that user will be unable to get resources from other users.

- In that proposal, the resource owner learns the relationships between the users who collaborate in the resource access. This represents a

privacy threat which would justify that some users might refuse collaboration. Nodes that refuse to collaborate cause other nodes to stay isolated.

In this chapter, we present a new protocol which offers the same features of [Carm07] and [Domi07] while addressing the drawbacks left open in [Domi07]. However, these shortcomings are not solved without cost: we assume the existence of an optimistic trusted third party (TTP) which only acts in case of conflict between the users of the social network. The optimistic TTP is not needed during the normal network execution. Therefore, we argue that this solution performs better than a (non-optimistic) TTP mediating all access requests.

Our scheme prevents the resource owner from learning the relationships and the trust levels between the users who collaborate in the resource access. In this way, the privacy threat detected in [Domi07] is solved and the number of users who might refuse collaboration due to privacy concerns is minimized. As a result, the chances for certain nodes to become isolated at certain periods of time are reduced.

The protocol we present uses multiplicative privacy homomorphisms. The needed cryptographic background is provided in Section 6.1. Section 6.2 presents our scheme in detail. This work has been published in [Domi08].

6.1 Multiplicative privacy homomorphisms

Privacy homomorphisms (PHs) are encryption transformations mapping a set of operations on cleartext to another set of operations on ciphertext. Basically, PHs are encryption functions $E : CT \rightarrow CT'$ allowing a set F' of operations on a ciphertext domain CT' to be carried out without knowledge

of the decryption function D. Knowledge of D allows the result of the corresponding set F of operations on a cleartext domain CT to be retrieved. A PH is called *multiplicative* when its set F of cleartext operations contains multiplication. A PH is called *probabilistic* if the encryption algorithm E involves some random mechanisms to choose the ciphertext corresponding to a given cleartext from a set of possible ciphertexts.

Privacy homomorphisms that will be used in our proposal below must be multiplicative, probabilistic and public-key. ElGamal [Elga85] is a probabilistic public-key cryptosystem of integers in the multiplicative group $\mathbb{Z}_p^*$, where p is a large prime. This cryptosystem has a multiplicative homomorphic property which fulfills all these requirements.

6.2 Homomorphic access control protocol for social networks

We follow the framework from [Carm07] modified according to [Domi07], that is, we consider that the node owning a resource rid (hereafter, the *resource owner*) establishes an access rule $AR = (rid, AC)$ where AC is the set of access conditions to be simultaneously satisfied to access rid. Several alternative access rules can be defined for a resource. An access condition is a tuple $ac = (v, rt, t_{min})$ where v is the resource owner. Such node must have a direct or indirect relationship with the node requesting resource rid (hereafter, the *requestor*), and rt, t_{min} are, respectively, the type and the minimum trust level that the relationship should have. The trust level t is a rational value such that $0 \leq t \leq 1$. We use a privacy homomorphism to encrypt the trust values contributed by the nodes in the social network. However, multiplicative homomorphisms are only available for integers in the

current literature. According to that, we propose to encode rational trust values as integer fractions; the details of the coding are given in Section 6.2.1.

Even though we use access rules and access conditions in a way similar to proposals [Carm07, Domi07], note that we differ from such schemes in that we do not use the maximum depth of the relationship as a requirement in access conditions. We have eliminated it because knowledge of the depth might be used by the resource owner to infer the trust level of the relationships between the users who collaborate in the resource access. We argue that this does not represent any security loss for the scheme since the minimum trust level and the type of the relationship are conditions that should be enough to decide whether a certain user can get access to a certain resource.

Each user U_i in the network owns two key pairs represented by (SK_i, PK_i) and (SSK_i, PSK_i). The former key pair corresponds to a public-key probabilistic multiplicative privacy homomorphism and it is used to encrypt/decrypt. The latter key pair is used to sign/verify. SK_i and SSK_i are the private keys. The corresponding public keys, PK_i and PSK_i, are assumed to be known and accepted by all users who have some interest in getting in touch with U_i. In the rest of this chapter, one-to-one communications are assumed to be confidential and authenticated (by properly using encryption and message authentication codes).

We agree with [Domi07] in that access should be enforced based on the relationship path between requestor and resource owner that yields the maximum trust level. This differs from the ideas presented in [Carm06, Carm07] where the trust level is computed taking into account all paths between requestor and resource owner, which might lead to overprotection: a requestor with a highly trusted direct relationship to the owner might be denied access just because there is also a requestor-owner indirect relationship with low

trust through a third user.

We next explain how to encrypt rational numbers used as trust levels by means of a homomorphism for integer values. Then, a simple version of our access control enforcement protocol is described in Section 6.2.2 which will help the reader to understand the new scheme. After that, the privacy problems that arise when using such a simple protocol will be discussed. Then, an enhanced solution will be described. Last but not least, it will be explained how the resource owner finally transmits his resource to the requestor (assuming the resource owner accepts the requestor).

6.2.1 Homomorphic encryption of rational values

As explained previously, users cannot encrypt rational numbers directly. According to that, we propose that users send fractions which will represent the real number linked to a certain trust value. As an example, a certain user U_i who wants to contribute a trust value of $0.\hat{3}$ will send fraction $1/3$ instead of the rational number. Note that the numerator and the denominator of the fraction are integers which can be encrypted in two different ciphertexts. Two different users U_1 and U_2 can multiply their own trust values (represented as fractions α_{U_1}/β_{U_1} and α_{U_2}/β_{U_2} respectively) in the ciphertext domain by performing the following operation:

$$\frac{E(\alpha_{U_1}) \otimes E(\alpha_{U_2})}{E(\beta_{U_1}) \otimes E(\beta_{U_2})} = \frac{E(\alpha_{U_1} \cdot \alpha_{U_2})}{E(\beta_{U_1} \cdot \beta_{U_2})}$$

where $E()$ denotes the encryption of a certain value following the privacy homomorphism in use and $\otimes$ denotes the ciphertext operation of such a cryptosystem corresponding to cleartext multiplication. At the end of the protocol, a user who is able to decrypt both resulting ciphertexts ($E(\alpha_{U_1} \cdot \alpha_{U_2})$

and $E(\beta_{U_1} \cdot \beta_{U_2})$) can divide the two recovered integer values to obtain the trust value as a rational number.

Nevertheless, this proposal introduces a privacy vulnerability: the user who is able to decrypt the ciphertexts may gain some clues from the resulting numerator and denominator on which fractions have been used to compute the final trust value. Since such fractions correspond to the trust values contributed by the users of the network, we argue that such information disclosure should be prevented. To address this situation we propose that, prior to encryption, each user U_i generates a random value ω_{U_i} that will be used to hide into a jumble of factors the numerator α_{U_i} and the denominator β_{U_i} that represent U_i's trust. Such hiding process is performed by multiplying numerator and denominator by ω_{U_i}:

$$\frac{E(\alpha_{U_i} \cdot \omega_{U_i})}{E(\beta_{U_i} \cdot \omega_{U_i})}$$

where ω_{U_i} is generated by multiplying a random sample (without replacement) of the prime numbers between 1 and a parameter n. Each prime is selected for the sample with probability γ. Each selected prime is raised to the power of an integer randomly selected between 1 and x (this integer is different for each prime).

Our scheme relies on multiplicative privacy homomorphisms to preserve the privacy of the users. Assuming that the privacy homomorphism in use is defined in the group $\mathbb{Z}_p^*$, the multiplication of all encrypted values must yield a result below p; otherwise information loss will occur. For the ciphertext containing the multiplication of all numerators this means

$$\prod_{i=1}^{s}(\alpha_{U_i} \cdot \omega_{U_i}) < p \tag{6.1}$$

where s is the total number of users whose trust is multiplied. An inequality analogous to Expression (6.1) can be written for the ciphertext containing the multiplication of all denominators. We next discuss which are the proper values for the parameters involved in those inequalities.

Parameter selection

As stated above, our scheme requires that the multiplication of encrypted numerators, resp. denominators, yields a result below p. The numerator α_{U_i} and the denominator β_{U_i} are the two elements of the fraction which represents U_i's trust value. Both elements are integers in the range $[0, \ldots, k]$. Value k is selected depending on the accuracy desired for the trust levels. As an example, with $k = 100$ we will guarantee an accuracy of two decimals: 0.93 can be represented by 93/100 or by 15/16 (numerators and denominators must be equal to or less than k). In what follows, $k = 100$ is taken.

Expression (6.1) depends on parameter s, which represents how many trust values will be multiplied together. In Section 6.2.6 we argue that usually $s \leq 6$. Nevertheless, we will fix it to $s = 7$ to leave enough room for situations with more trust levels to be multiplied.

In terms of length, the worst case for ω occurs when all prime numbers in $[1, \ldots, n]$ are selected, and all of them are raised to the power of x (the maximum value). In order to compute the maximum length, we assume that all n numbers are selected (prime or not). According to that, $(n!)^x$ is an upper bound on ω. It means that $\prod_{i=1}^{s}(\alpha_{U_i} \cdot \omega_{U_i})$ will be at most $(k \cdot (n!)^x)^s$. The length of this expression is:

$$|(k \cdot (n!)^x)^s| = s \cdot \log_2(k \cdot (n!)^x) = s \cdot (\log_2(k) + x \cdot \log_2(n!)) \qquad (6.2)$$

If we set the length of p to 1024 bits, Expression (6.2) must stay below 1024 bits. Given $k = 100$ and $s = 7$, Table 6.1 shows values of n and x such that this requirement holds. It can be observed that for $x = 10$ (which is high enough), value n can be taken up to 7. In addition to that, if we set $n = 11$ (which is high enough too), x can be increased up to 5. These results prove that the proposed method for real number encoding is feasible and works properly.

Note that the considered upper bound on ω is a loose one: all numbers (prime or not) in $1, \ldots, n$ are selected. An average length for ω considering only the prime numbers can be computed as

$$\sum_{\forall \text{ prime } i \in (1,\ldots,n)} \gamma \cdot x/2 \cdot \log_2(i)$$

Table 6.1: For $k = 100$ and $s = 7$, upper bound given by Expression (6.2) for different values of n and x

n	$s \cdot (\log_2(k) + x \cdot \log_2(n!))$				
	$x = 3$	$x = 5$	$x = 10$	$x = 15$	$x = 20$
3	100.79	136.98	227.45	317.93	408.40
5	191.55	288.25	529.99	771.73	1013.47
7	304.79	476.98	907.45	1337.92	1768.40
11	576.77	930.27	1460.53	2697.81	3581.58
13	729.76	1185.26	2324.02	3462.78	4601.53
16	975.76	1595.26	3144.02	4692.77	6241.53
17	1061.60	1738.32	3430.14	5121.96	6813.77

6.2.2 Simple homomorphic protocol

We present in this subsection a first protocol using homomorphic encryption to guarantee relationship privacy.

Protocol 1 (Simple homomorphic protocol)

1. *The resource owner B advertises to the network a certain resource rid he wants to share. Such an advertisement is signed by B, and contains all the access rules $AR_1, \ldots, AR_r$ defined for rid. That is, $\{AR_1, \ldots, AR_r\}_{SSK_B}$.*

2. *The requestor A is interested in rid. In order to gain access to that resource, A sees to it that the resource owner receives one or several relationship certificates proving that the requestor satisfies all access conditions corresponding to at least one of the access rules. Several cases can be distinguished depending on the relationship depth between the requestor and the resource owner:*

   (a) Depth 1. *A and B have a direct relationship, that is, A is related to B through a relationship of type rt and trust level t_{AB} represented by a tuple (rt, t_{AB}) and B is related to A through a relationship (rt, t_{BA}). Note that the relevant trust level here is t_{BA} (how much B trusts A) which is assumed to be* unknown *to A. In this case A directly asks B whether he is granted access to the resource on the basis of (rt, t_{BA}). If B evaluates that rt and t_{BA} satisfy the set of access conditions targeted by A, then A is granted access. Otherwise, A is required to resort to other direct relationships or indirect relationships.*

   (b) Depth 2. *If A and B have no direct relationships (or these are not enough to buy access to A) then A asks to all users with whom A is directly related whether they have direct relationships of the relevant type rt with B. Assume C is directly related to both A*

and B with relationship type rt. Then C sends to B a pair of messages $PK_B(rt), PK_B(t_{CA})$ encrypted under the public key of the resource owner B.

The trust values are rational values homomorphically encrypted as fractions as detailed in Section 6.2.1. According to that, $PK_B(t_{CA})$ corresponds to:

$$PK_B(t_{CA}) \Leftrightarrow \frac{PK_B(\alpha_{CA} \cdot \omega_{CA})}{PK_B(\beta_{CA} \cdot \omega_{CA})}$$

However, for the sake of readability, we prefer to write $PK_B(t_{CA})$ instead of the above fraction. Note that the encryption of the relationship type rt is simpler, because it is not a rational number.

After C has sent to B the pair $PK_B(rt), PK_B(t_{CA})$, C tells A that a message was sent to B, but does not reveal *its content. At this point B evaluates whether a relationship of type rt and trust level $t_{CA} \cdot t_{BC}$ is enough to grant access of A to rid.*

(c) Depth 3. *If access at depth less than or equal to 2 cannot be obtained then A requests to users C directly related to him to attempt access with depth 2 on A's behalf: each C directly related to A contacts his other directly related users D about possible direct relationships between D and B (similarly to what A did in Step 2b). If a D with direct relationships to C and B exists, D must multiply his own trust value related to C times the current trust value which comes from C. To do that, D computes:*

$$PK_B(t_{CA} \cdot t_{DC}) = PK_B(t_{CA}) \otimes PK_B(t_{DC})$$

where $\otimes$ denotes the ciphertext operation of the privacy homomorphism corresponding to cleartext multiplication. According to the homomorphic fraction coding described in Section 6.2.1, $PK_B(t_{CA})\otimes PK_B(t_{DC})$ corresponds to:

$$PK_B(t_{CA}) \otimes PK_B(t_{DC}) \Leftrightarrow \frac{PK_B(\alpha_{CA} \cdot \omega_{CA}) \otimes PK_B(\alpha_{DC} \cdot \omega_{DC})}{PK_B(\beta_{CA} \cdot \omega_{CA}) \otimes PK_B(\beta_{DC} \cdot \omega_{DC})}$$

The former expression implies the following correspondence:

$$PK_B(t_{CA} \cdot t_{DC}) \Leftrightarrow \frac{PK_B(\alpha_{CA} \cdot \omega_{CA} \cdot \alpha_{DC} \cdot \omega_{DC})}{PK_B(\beta_{CA} \cdot \omega_{CA} \cdot \beta_{DC} \cdot \omega_{DC})}$$

Finally, D sends a message containing $(PK_B(rt), PK_B(t_{CA} \cdot t_{DC}))$ to B. Upon receiving this message, B evaluates whether a relationship of type rt and trust level $t_{CA} \cdot t_{DC} \cdot t_{BD}$ is enough to grant access of A to rid. Note that B receives the product $t_{CA} \cdot t_{DC}$, but he cannot discover the individual trust levels which have been multiplied.

Figure 6.1 represents a path between the requestor A and the resource owner B through the social network. The picture shows the encrypted trust value as computed on its way from A towards B. The ciphertext containing the relationship type (rt) has been omitted.

(d) Successive depths. *In case of failure at depth 3, successive depths are tried in a similar way.*

Remark. When the resource owner advertises the access rules for a resource, the access conditions in those rules leak the relationships the owner is involved in (*e.g.* if the owner accepts rt = 'Colleague at company X' this

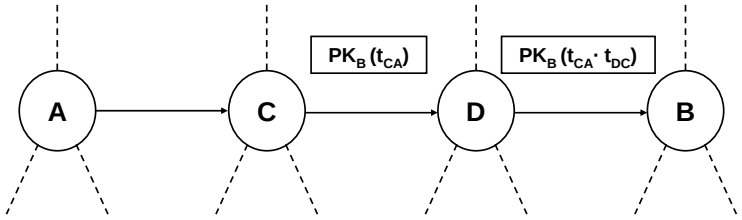


Figure 6.1: Resource request in the simple homomorphic protocol

means that he works at Company X). In [Carm07] the relationship type is kept confidential through a rather complex symmetric encryption scheme. A first problem of this scheme arises when the same relationship type is encrypted using two different keys by two different user communities and these merge at a later stage; another, perhaps more serious problem is how to revoke the key used to encrypt a given relationship type. An alternative and simpler strategy is to "camouflage" the real relationship types among a large number of bogus relationships; then access conditions are published some of which use real relationship types and most of which use bogus relationship types. A bogus relationship type rt' is one that has never been established by the owner with anyone, so that no one can request access based on rt'. The advantage is that a snooper cannot tell bogus relationships from real ones, so that he does not know which relationships the owner is actually involved in.

6.2.3 Anonymous homomorphic protocol

Protocol 1 above is unsafe if nodes are not anonymous and there is a relationship of depth 2 between requestor and owner. We next explain why. Let A be a requestor, C be an intermediate node and B be a resource owner. During the protocol B gets $PK_B(t_{CA})$ and decrypts it to obtain t_{CA}, that is,

the trust level assigned by C to A. This represents an unavoidable privacy problem in Protocol 1, as B needs to compute $t_{CA} \cdot t_{BC}$ in order to evaluate whether A is trusted enough to access a resource. Of course, one might argue that B does not know how many nodes there are between C and A (there might be several, and t_{CA} might be the product of the trust levels between C and A); however, this seems a rather weak protection, because nodes are not anonymous. Worse yet, the above weakness can easily spread. Once t_{CA} is learnt by B, for any path of length 3 D-A-C-B, then B will learn t_{AD}. And so on.

To avoid that, we propose to enforce anonymity for nodes which are not directly related. For example, in a path of length 3 A-C-D-B, this kind of anonymity means that B only knows D and he has no knowledge about which nodes or how many nodes are behind D; also, D only knows C, but not A. Furthermore, also for the sake of anonymity, C should not know that A is the requestor; in order to make himself undistinguishable from an intermediate node, A sends to C an encrypted trust value, which in A's case must be an encryption of the initial neutral trust 1, that is $PK_B(1)$ (note that, since we are assuming the use of probabilistic homomorphic encryption, C cannot discover that $PK_B(1)$ is an encryption of the maximum trust 1).

As mentioned above, node anonymity was previously proposed in [Wang06]. However, node re-identification in that scheme was still possible because relationship types and trust levels were public. In our scheme, we combine node anonymity with encryption of trust levels, in order to provide real privacy for the users of the social network.

A problem which arises is that dishonest users may take advantage of anonymity to disrupt the system without being punished. We propose a *liability mechanism* to thwart such disruption.

Liability mechanism

In our scheme, anonymity exists in indirect relationships (with depth greater than one), but direct relationships are non-anonymous: each node knows his directly related nodes. Furthermore, we require that every pair of directly connected nodes A and C should not only know but authenticate each other before engaging in any protocol transaction between them. As a result, a dishonest user E cannot impersonate a certain node of the social network (for example node A) and she cannot repudiate being user E.

According to that, an intermediate node C directly related to (and thus authenticated by) a resource owner B can be held liable by B for any harm that results from the encrypted trust level that C forwards to B. In turn, C can extend this liability to his direct relationships along the requestor-owner path. This liability transmission is the same used in chained subcontracting in daily life (*e.g.* the first subcontractor is liable in front of the main contractor, the second subcontractor is liable in front of the first subcontractor and so on).

Now, let us imagine that a resource owner B has followed the protocol and he has given an anonymous node access to a certain resource (*e.g.* a movie). Later on B discovers that this resource is being unlawfully re-distributed over the Internet. Following our liability mechanism, B will point C as guilty since C is the only user in the requestor-owner path known by B. Then B will take proper countermeasures against C. In turn, C (if he is not the dishonest user) will do the same with the direct node he knows and so on up to the requestor.

This mechanism has a major problem: at the end of the process, all users in the path pay the consequences of the misbehavior by a single node. Therefore, this situation can discourage users from collaborating in resource

access. To solve that, we add to our liability mechanism the use of certificates as a proof of the direct relationship between two users in the access to a certain resource. For example, in a path of length 3 *A*-*C*-*D*-*B*, user A will give a certificate $Cert_{AC}$ to C as a proof of their direct relationship in a certain resource access. In turn, C gives a certificate $Cert_{CD}$ to D and so on. A certificate for a relationship where node N_1 forwards the encrypted trust level to node N_2 is constructed as follows:

$$Cert_{N_1N_2} \leftarrow \{id||N_1||N_2||K_{N_1}(t_{N_1N_0})||time_stamp\}_{SSK_{N_1}} \tag{6.3}$$

In Expression (6.3), *id* is the identifier assigned to a certain resource access (which is linked to a certain item); N_1 is the identity of the node who constructs and sends the certificate; N_2 is the receiver; N_0 is the node preceding N_1 and $K_{N_1}(t_{N_1N_0})$ is the trust value that N_1 assigns to N_0 (encrypted under a secret key K_{N_1} only known by N_1). Note that when N_1 is the requestor, N_0 does not exist. In this case $t_{N_1N_0}$ is taken to be 1, the neutral value for multiplication. The certificate is signed by N_1 and contains a *time_stamp* which reflects when it was generated. The public key PSK_{N_1} for verification is known and accepted as valid by all nodes in the network.

A user U who receives a certificate must store it in a safe place. Later on, if someone in the relationship path misbehaves, U can use the certificate to prove his innocence. We next explain how these certificates are used to protect the system against a misbehaving node (whether it is a requestor or an intermediate node).

Figure 6.2 represents the same situation shown in Figure 6.1 but following our proposed anonymous homomorphic protocol. Figure 6.2 shows a resource request from the point of view of the resource owner (node B). In this situation, B only knows that the computed trust value contains a trust from

D ($t_{D?}$) but he does not know who is behind D. Actually, B does not know how many nodes are behind D neither who is the requestor.

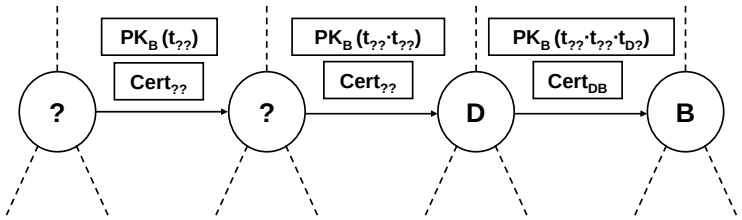


Figure 6.2: Resource request in the anonymous homomorphic protocol (from the point of view of the resource owner)

Regarding the computational cost, our anonymous homomorphic protocol using the proposed liability mechanism requires intermediate nodes involved in a protocol execution to perform the following operations:

1. Use the privacy homomorphism to encrypt a trust value t.

2. Compute the ciphertext operation of the privacy homomorphism on the received encrypted trust value t' and the own encrypted trust value t.

3. Generate the certificate $Cert_{N_1N_2}$ required by the liability mechanism.

A resource requestor only executes operations (1) and (3). A resource owner only decrypts the received ciphertexts containing the final trust value and the relationship type. Later, the resource owner checks the set of access conditions in plain text. Note that all these operations are feasible in full-fledged computers (*e.g.* desktop computers, laptops, powerful mobile devices, etc.). The social network that we envisage runs on this kind of devices so we argue that our protocol will perform properly in terms of computation when deployed in a real environment.

Protecting the system by partial revocable anonymity

Nodes in the relationship path can misbehave in two different ways:

- A *requestor* who has gained access to a certain resource can later unlawfully re-distribute it over the Internet.
- An *intermediate node* can contribute false trust levels to the relationship path.

If the requestor misbehaves, an intermediate node U can use the certificate he owns to prove to the resource owner that he is only an intermediate node who has forwarded the encrypted trust level. In this way, node anonymity vanishes, because each intermediate node publishes his direct relationship in the path until the requestor is reached, who will be properly punished for his misbehavior. This procedure has a major problem: a dishonest resource owner can falsely pretend that a certain requestor has misbehaved in order to gain knowledge of all the relationships in the owner-requestor path.

We deal with this situation by proposing partially rather than totally revocable anonymity. According to that, we assume the existence of an *optimistic trusted third party* who is able to access the certificates to judge whether someone is misbehaving. In this way, the resource owner does not get any information on the requestor-owner path. An optimistic TTP is a trusted authority who only acts in case of conflict between the users of the social network. It is not needed during the normal network operation, which is much more efficient than requiring a (non-optimistic) trusted authority to mediate all transactions.

We now turn to the second type of misbehavior, in which an intermediate node contributes false trust levels to the relationship path. Even though trust values are kept secret, a certain user A having a direct relationship

with another user B can make a good guess about the value of t_{BA}. There are two cases:

- If A relies on B's help to access resources and B misbehaves by contributing a fake t_{BA} *lower* than the real one, A will end up suffering a DoS attack. We believe that this situation should be avoided, or at least limited to the extent possible. Precisely to that end, Expression (6.3) includes the trust value $t_{N_1N_0}$ contributed by the node N_1 generating $Cert_{N_1N_2}$. Since $t_{N_1N_0}$ is encrypted under K_{N_1}, nobody but N_1 can recover $t_{N_1N_0}$. However, if a user thinks that he is facing a DoS attack from another user, he can report this situation to the optimistic TTP who requests the secret key used by each intermediate node; then, the optimistic TTP checks the trust values contributed by each intermediate node and takes action against those behaving dishonestly. Note that no one but this authority will know the trust value associated to each direct relationship.

- If an intermediate node contributes a fake trust level *higher* than the original one, this can result in access being granted to a non-deserving requestor. Beyond illegal access, other problems may arise if the requestor unlawfully re-distributes the accessed resource or abuses it in other ways; see subsection above (which details the liability mechanism) for a description of the defenses against a misbehaving requestor.

6.2.4 Accessing the resource

Our scheme is based on preserving node anonymity as long as nodes behave honestly. For that reason, a resource owner B who accepts to send a certain resource rid to a requestor D cannot transmit rid directly to D because the

identity of the requestor is unknown to B. Instead of that, the resource must follow in reverse order the same relationship path which was previously used to decide on the requestor's access.

If this store-and-forward process involves too much bandwidth consumption for intermediate nodes (*e.g.* if the resource is a movie or some kind of large item), an alternative is for the requestor to send together with his query an ftp address where the resource owner can upload the resource. Note that this option implicitly assumes the existence of a trusted third party: by default, ftp downloading is not anonymous and in this case the ftp site must be trusted to preserve the anonymity of all requestors who use it; if some anonymizer is used between the requestor and the ftp site, then this anonymizer is playing the role of a TTP who should preserve the anonymity of the requestor. Whatever the case, since the TTP can be totally external and unrelated to the social network or its users, assuming that such an entity will keep secret the identity of the requestor seems plausible.

6.2.5 Security and privacy analysis

We next explain the assumed adversary model and the possible attacks the system has to be robust against. Later, we detail the protocol behavior against each considered attack.

Adversary model

Our attacker model assumes that the adversary is a node in the social network. Such an adversary can collude with others to attack the system. We consider that the computational power of an attacker does not permit him to break current computationally secure cryptosystems.

We divide the possible attacks in four categories. The first three categories refer to the role adopted by the attacker: *intermediate node*, *resource requestor* and *resource owner*. The last category refers to *collusions between nodes*. We next list the considered attacks within each category.

- *Intermediate node.*
  - Learn the trust levels of the previous nodes in the requestor-owner path.
  - Alter the received trust or contribute a fake one.
  - Refuse to collaborate.
- *Resource requestor.*
  - Increase the trust sent by other users in order to get access to a certain resource.
- *Resource owner.*
  - Learn the trust levels of the nodes in the requestor-owner path.
- *Collusion between nodes.*
  - Learn the trust level of an honest user surrounded by colluders.

Protocol behavior against the considered attacks

Learn the trust levels of the nodes in the requestor-owner path. This attack can be performed by an intermediate node or by the resource owner. The resource owner is able to decrypt the encrypted product of trust levels and get the final trust value. However, the owner cannot learn from that value which nodes have collaborated nor the individual trust

level contributed by each of them. An intermediate node is less dangerous than the resource owner, because he does not know the secret key needed to decrypt the received product of trust levels.

Alter the received trust or contribute a fake one. As said above, an intermediate node can not learn the computed trust value received from his "upstream" neighbor (the neighbor previous to him in the requestor-owner path) because it is encrypted. However, the intermediate node can replace the received trust value with any trust value he desires and such a behavior will go undetected.

As explained in Section 6.2.3, our proposal uses certificates which contain the trust values contributed by each node. In addition to that, we assume the existence of an optimistic TTP who is able to check such values and take action against dishonest nodes. We argue that both measures will discourage users from performing this attack.

Refuse collaboration. An intermediate node can decide not to collaborate in a resource access. There is no defense against this. However, in our scheme there is no price to be paid by the intermediate nodes who collaborate: the privacy toll in [Domi07] (mentioned in the introduction) is solved by our scheme and, with the use of an ftp service, an intermediate node is not even required to spend any of his bandwidth to enable resource access.

Therefore, intermediate nodes do not have any objective reason to refuse collaboration. In fact, there is even an incentive for them to collaborate: a node who routinely refuses collaboration will be in a bad position when he later seeks collaboration as a resource requestor.

Unlawful trust increase. The resource requestor cannot unlawfully increase

any trust level in the requestor-owner path because he does not see any of the relationship messages exchanged between intermediate nodes that will be used by the resource owner to decide whether the requestor is granted access.

Collusion between nodes. Collusion between nodes is successful when adversary nodes are surrounding the victim and the resource owner is one of the colluding nodes. In this situation the colluders learn the trust value which the victim has assigned to his adversary upstream neighbor. However, we argue that the surrounded victim can decide not to collaborate if he does not trust his neighbors (upstream or downstream) or the owner (the identity of the latter is public, since he is offering a resource).

6.2.6 Simulation

We have simulated our protocol in a realistic environment to observe its performance. The network simulator *ns-2* [Netw08] was used for this purpose.

Central to our proposal is the availability of on-line users who have relationships between them and enable a certain requestor to get access to a certain resource. Therefore, the lack of active users in the network is a potential problem. After pondering this issue, we decided to check the impact of the shortage of on-line users in small networks. For large networks with lots of connected users, the problem is less likely. With this in mind, the proposed scheme was tested in four social networks consisting, respectively, of 100, 300, 500 and 1000 users.

In each social network, the connection topology between users and their associated relationships were fixed as follows before starting the simulation:

- Each user was connected to (*i.e.* held relationships with) a number of users ranging between 1 and 30. The precise figure was decided using the *power-law distribution*. As stated in [Libe05], typical social networks are reasonably well approximated using this distribution. As a result, in our simulated social network 78.76% of the users held a number of connections ranging between 1 and 15. Also, 31.07% of the users held only 1 or 2 connections.

- Each node could establish three different types of relationship with other nodes. Up to one relationship of each type was allowed between each pair of nodes, so that there could be up to three relationships between two nodes. Whether there existed a relationship of a certain type between two nodes was uniformly randomly decided.

- The trust level assigned to each existing relationship was randomly and uniformly chosen in the range $[0.5, 1]$.

A simulation test consisted of running a request by a requestor to access a resource advertised by an owner. Both requestor and owner were randomly chosen. The simulation ended when:

- either the requestor was granted access to the resource, which happened if the resource owner could compute a trust value greater than or equal to 0.4;

- or a time-out occurred (this happened when either no path existed between requestor and owner or the existing paths yielded a computed trust less than 0.4).

For each social network, several fractions of simultaneous on-line users were considered: 10%, 30%, 50%, 70% and 90%. For each fraction, 50 tests

were run and the average results over the 50 tests were computed. The results for the four social networks considered are shown in Tables 6.2, 6.3, 6.4 and 6.5. The contents of those tables are commented below.

Table 6.2: Results for a social network of 100 users

#on-line nodes	average access prob.	average # interm. nodes	average #messages	average trust level
10	0.09	2.89	192.91	0.42
30	0.20	3.10	886.72	0.48
50	0.36	2.94	1221.30	0.59
70	0.50	3.44	1635.24	0.50
90	0.70	3.29	1825.60	0.46

Table 6.3: Results for a social network of 300 users

#on-line nodes	average access prob.	average # interm. nodes	average #messages	average trust level
30 (10%)	0.11	3.33	260.00	0.50
90 (30%)	0.24	3.92	1298.25	0.49
150 (50%)	0.46	4.09	2493.65	0.45
210 (70%)	0.56	4.07	3009.82	0.42
270 (90%)	0.73	4.03	3616.33	0.44

Simulation results

For each social network and each fraction of on-line nodes (*# on-line nodes*), the following results are reported:

- *Average access probability.* This is the average probability that a certain access request is granted (*i.e.* that a computed trust greater than or equal to 0.4 is obtained).

Table 6.4: Results for a social network of 500 users

#on-line nodes	average access prob.	average # interm. nodes	average #messages	average trust level
50 (10%)	0.14	4.00	683.75	0.46
150 (30%)	0.28	3.43	1353.14	0.62
250 (50%)	0.49	5.36	2163.32	0.41
350 (70%)	0.66	5.33	3593.15	0.43
450 (90%)	0.82	4.59	3771.22	0.41

Table 6.5: Results for a social network of 1000 users

#on-line nodes	average access prob.	average # interm. nodes	average #messages	average trust level
100 (10%)	0.18	5.20	815.37	0.41
300 (30%)	0.32	4.92	2880.25	0.46
500 (50%)	0.52	5.67	3373.04	0.42
700 (70%)	0.71	5.47	4412.25	0.43
900 (90%)	0.89	4.71	4628.02	0.42

- *Average number of intermediate nodes.* This is the average number of intermediate nodes in the path between the requestor and the resource owner.

- *Average number of messages.* This is the average total number of messages generated by the nodes in the social network until the launched resource request is granted or a time-out occurs.

- *Average trust level.* This is the average trust level for the accepted requests.

From the results in Tables 6.2, 6.3, 6.4 and 6.5, it can be observed that better performance is obtained when there is a higher fraction of on-line

users: the average access probability is higher and the requestor-owner path is shorter. This is what one would expect.

It can be seen in those tables that when 30% or less users are on-line it is difficult to get access to a resource. When 50% or more users are on-line the system works reasonably well. Note that these results are obtained in the worst possible scenario for our protocol: relatively small networks where a big proportion of the users have no more than two, randomly established connections. Under these circumstances, it is pretty likely that several nodes find themselves isolated (without connections to other on-line nodes). In a more realistic scenario with several human user communities with common interests, it would be more likely for nodes in the same community to be on-line at the same time and hold more connections between them. This similarity of habits would render node isolation rarer, even if there is a low fraction of on-line nodes.

Also, it can be observed that, whatever the proportion of on-line nodes, the average number of intermediate nodes in a requestor-owner path is no more than six. This should not be surprising, because social networks are affected by the "six degrees of separation" phenomenon [Milg67]: long ago, S. Milgram experimentally showed that any two people in the United States are connected through about six intermediate acquaintances, implying that we live in a rather small world. According to the "six degrees of separation" concept, for huge social networks the separation between any two users will still be a maximum of six intermediate nodes. This is good news for our scheme in the sense that the number of generated messages until a resource request is accepted does not grow linearly with the number of on-line nodes. In fact, we can observe in Table 6.5 that the total number of messages sent is quite similar when there are 300 and 900 on-line nodes. These results show

that our proposal is scalable enough to work properly when deployed in real social networks.

Chapter 7

Conclusions

7.1 Concluding remarks

In this thesis, we have pointed out the importance of providing security and privacy for new emergent applications based on special-purpose networks. More specifically, we have covered different security and privacy issues related to some applications based on four types of special-purpose networks. These are:

- Secure information transmission in many-to-one scenarios with resource-constrained devices such as sensor networks.
- Secure and private information sharing in mobile ad hoc networks (MANETs).
- Secure and private information spread in vehicular ad hoc networks (VANETs).
- Private resource access in social networks.

The primary concern has been to offer a broad overview of current techniques for providing security and privacy in each environment.

Regarding many-to-one communications for resource-constrained devices, we have focused on secure many-to-one lossless transmission. More precisely, we have presented two new proposals following this paradigm.

Different models for information sharing in MANETs have been studied. We have presented two new schemes dealing with security and privacy issues in this environment.

How to provide trustworthy information spread over VANETs has been studied in detail. More precisely, we have studied both *a priori* and *a posteri* countermeasures against fake messages from internal attackers. We have presented a system that relies on *a priori* countermeasures and provides secure vehicle-generated announcements on VANETs. The new system has been compared with current proposals in the literature.

Last but not least, we have addressed how to provide resource access in social networks while preserving the privacy of the users. A new proposal has been presented.

7.2 Results of this thesis

We summarize here the results presented in this thesis.

In Chapter 3, our contributions regarding efficient and secure many-to-one symbol transmission have been presented. First, a new protocol that provides an optimal message length has been proposed. The new protocol works properly in environments where bandwidth is scarce. This protocol uses multisignatures and offers the four basic security properties: confidentiality, integrity, authentication and non-repudiation. Immediate detection

of corrupted messages is provided too. Next, the first scheme in the literature that offers secure many-to-one symbol transmission for sensor networks has been proposed. This protocol also provides an optimal message length and it is computationally suitable for resource-constrained devices, which are quite common in sensor networks. We have referred to the computational capabilities of real sensor devices to prove the deployability of this proposal in real environments. This scheme achieves the following security properties: confidentiality, authentication and integrity. Detection of corrupted messages is performed using an *a posteriori* tracing algorithm.

Chapter 4 contains two contributions about information sharing in mobile ad hoc networks. Both of them provide security and privacy for the users of the network. Incentives are given to avoid user misbehavior. The first construction provides information in a urban environment. A certain user, regardless of her location, can request information any time using her mobile device and its wireless connection. This system provides incentives to encourage users to become distributed information servers. Regarding the second contribution, it is a new scheme designed to disseminate advertisements through mobile ad hoc networks. This proposal outperforms the current proposals in literature by offering security and privacy without requiring the participation of any trusted third party (except for a certification authority that certifies the merchant's public key). In addition to that, we propose a new approach to reward nodes that collaborate in the dissemination according to how long they have been holding an advertisement. This proposal does not bound the number of transfers for an advertisement (and thus its spreading range) and rewards collaborative nodes with e-coins proportionally to their task.

Chapter 5 presents a new system that provides secure vehicle-generated

announcements on VANETs. This scheme relies on *a priori* measures against internal attackers (vehicles in the VANET sending fake messages). Thanks to the use of threshold signatures, our system outperforms previous proposals in message length and computational cost. Regarding privacy, three different variants of the system have been proposed to achieve privacy without losing trustworthiness: the first variant is a special case of the second one and is better suited to dense VANETs, whereas the second and third variants can be used as fallbacks for sparse VANETs. The feasibility of this scheme has been studied using simulations.

The last chapter of this thesis presents a privacy-preserving resource access protocol for social networks. This new protocol achieves protection of relationship privacy, with the advantage of being fault-tolerant and free of mediating TTPs (although an optimistic TTP is used in case of conflict). On the whole, our scheme offers the same features as [Carm07] and [Domi07] while addressing the functionality and privacy drawbacks of those previous protocols. The simulated performance of our proposal shows that it can be successfully deployed in real environments because it is scalable and it provides reasonable resource availability.

7.3 Future research

We sketch here some open problems that remain to be solved and possible extensions to some of the presented contributions that will be addressed in the future.

In the field of many-to-one communications, our future research will be directed to design schemes for secure many-to-one *lossy* transmission. More specifically, our objective is to design a scalable protocol that allows the

base station to get the result of some mathematical function (*e.g.* Min/Max, average computation...) applied to the data stored in the leaves. Such a system should provide confidentiality, authentication and integrity to the data transmitted from the leaves towards the root.

More research on vehicular ad hoc networks is planned. Our future work in this field will be directed to the construction of a scheme that relies on *a posteriori* measures against internal attackers.

Regarding private resource access protocols for social networks, the presented scheme should be extended to eliminate the need for an optimistic TTP.

Our contributions

[Cast07] J. Castellà-Roca, V. Daza, J. Domingo-Ferrer, J. Manjón, F. Sebé and A. Viejo, "An incentive-based system for information providers over peer-to-peer mobile ad-hoc networks", *Lecture Notes in Computer Science*, Vol. 4617, pp. 380–392, 2007.

[Daza08] V. Daza, J. Domingo-Ferrer, F. Sebé and A. Viejo, "Trustworthy Privacy-Preserving Car-Generated Announcements in Vehicular Ad Hoc Networks", *IEEE Transactions on Vehicular Technology*, (to appear 2008).

[Domi08] J. Domingo-Ferrer, A. Viejo, F. Sebé and Úrsula González-Nicolás, "Privacy Homomorphisms for Social Networks with Private Relationships", *Computer Networks*, (to appear 2008).

[Sebe07b] F. Sebé, A. Viejo and J. Domingo-Ferrer, "Secure Many-to-One Symbol transmission for Implementation on Smart Cards", *Computer Networks*, vol. 51, no. 9, pp. 2299–2307, 2007.

[Viej07] A. Viejo, F. Sebé and J. Domingo-Ferrer, "Secure and private incentive-based advertisement dissemination in mobile ad hoc networks", *Lecture Notes in Computer Science*, vol. 4752, pp. 185–198, 2007.

[Viej08] A. Viejo, F. Sebé and J. Domingo-Ferrer, "Secure and scalable many-to-one symbol transmission for sensor networks", *Computer Communications*, vol. 31, no. 10, pp. 2408–2413, 2008.

Bibliography

[Armk07] F. Armknecht, A. Festag, D. Westhoff and K. Zeng, "Cross-layer privacy enhancement and non-repudiation in vehicular communication", *4th Workshop on Mobile Ad-Hoc Networks (WMAN)*, 2007.

[Ashr06] R. Ashri, S. D. Ramchurn, J. Sabater, M. Luck and N. R. Jennings, "Trust evaluation through relationship analysis", *4th International Joint Conference on Autonomous Agents and Multiagent Systems*, ACM, pp. 1005–1011, 2005.

[Barn06] S. B. Barnes, "A privacy paradox: social networking in the United States", *First Monday*, vol. 11, no. 9, 2006.

[Barr02] P. S. L. M. Barreto, H. Y. Kim, B. Lynn, and M. Scott, "Efficient algorithms for pairing-based cryptosystems", *Lecture Notes in Computer Science*, vol. 2442, pp. 354–368, 2002.

[Bene05] Z. Benenson, "Authenticated Queries in Sensor Networks", *Lecture Notes in Computer Science*, vol. 3813, pp. 54–67, 2005.

[Berg07] I. Berger, "Standards for car talk", *IEEE The Institute*, vol. 31, no. 1, 2007.

[Berk04] University of California Berkeley, "Tiny OS Hardware Designs", 2004. `http://www.tinyos.net/scoop/special/hardware`

[BlaB05] E.-O. Blaß and M. Zitterbart, "Towards Acceptable Public-Key Encryption in Sensor Networks", *ACM 2nd International Workshop on Ubiquitous Computing*, pp. 88–93, 2005.

[Blun97] C. Blundo and D. R. Stinson, "Anonymous secret sharing schemes", *Discrete Applied Mathematics*, vol. 77, pp. 13–28, 1997.

[Blun05] C. Blundo, S. Cimato and A. De Bonis, "Secure E-Coupons", *Electronic Commerce Research Journal*, vol. 5, no. 1, pp. 117–139, 2005.

[Bold03] A. Boldyreva, "Efficient threshold signatures, multisignatures and blind signatures based on the Gap-Diffie-Hellman-group signature scheme", *Lecture Notes in Computer Science*, vol. 2567, pp. 31–46, 2003.

[Bone01] D. Boneh, B. Lynn and H. Shacham, "Short signatures from the Weil pairing", *Lecture Notes in Computer Science*, vol. 2248, pp. 514–532, 2001.

[Butt00] L. Buttyan, J. Hubaux, "Enforcing service availability in mobile ad-hoc WANs", *Proceedings of the 1st ACM MobiHoc*, pp. 87–96, 2000.

[Butt03] L. Buttyan and J. Hubaux, "Stimulating Cooperation in Selforganizing Mobile Ad Hoc Networks", *ACM/Kluwer Mobile Networks and Applications*, vol. 8, no. 5, 2003.

[C2cc08] CAR 2 CAR Communication Consortium, 2008.
`http://www.car-2-car.org`

[Carm06] B. Carminati, E. Ferrari and A. Perego, "Rule-based access control for social networks", *Lecture Notes in Computer Science*, vol. 4278, pp. 1734–1744, 2006.

[Carm07] B. Carminati, E. Ferrari and A. Perego, "Private relationships in social networks", *Private Data Management*, IEEE Press, 2007.

[Cast05] C. Castelluccia, S. Jarecki, J. Kim and G.Tsudik, "Secure acknowledgment aggregation and multisignatures with limited robustness", *Computer Networks*, vol. 50, no. 10, pp. 1639–1652, 2006.

[Chau89] D. Chaum, "Privacy Protected Payments: Unconditional Payer and/or Payee Anonymity", *Smart Card 2000*, pages 69–92. 1989.

[Chau90] D. Chaum, B. Den Boer, E. Van Heyst, S. Mjolsnes and A. Steenbeek, "Efficient offline electronic checks (extended abstract)", *Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques on Advances in Cryptology*, 1990.

[Damg01] I. Damgård and M. Koprowski, "Practical threshold RSA signatures without a trusted dealer", *Lecture Notes in Computer Science*, vol. 2045, pp. 152–165, 2001.

[Dimi05] T. Dimitriou, Efficient Mechanisms for Secure Inter-node and Aggregation Processing in Sensor Networks, *Lecture Notes in Computer Science*, vol. 3738, pp. 18–31, 2005.

[Dimi06] T. Dimitriou, Securing Communication Trees in Sensor Networks, *Lecture Notes in Computer Science*, vol. 4240, pp. 47–58, 2006.

[Djen05] D. Djenouri, L. Khelladi, A.N. Badache, A survey of security issues in mobile ad hoc and sensor networks, *IEEE Communications Surveys & Tutorials*, vol. 7, num. 4, pp. 2–28, 2005.

[Domi04] J. Domingo-Ferrer, A. Martínez-Ballesté and F. Sebé, "Secure reverse communications in a multicast tree", *Lecture Notes in Computer Science*, vol. 3042, pp. 807–816, 2004.

[Domi07] J. Domingo-Ferrer, "A public-key protocol for social networks with private relationships", *Modeling Decisions for Artificial Intelligence*, *Lecture Notes in Computer Science*, vol. 4617, pp. 373–379, 2007.

[Dötz06] F. Dötzer, "Privacy issues in vehicular ad hoc networks", *Lecture Notes in Computer Science*, vol. 3856, pp. 197–209, 2006.

[Ecds99] ANSI X9.62-1998, "Public Key Cryptography for the Financial Services Industry: The Elliptic Curve Digital Signature Algorithm (ECDSA)", American National Standard for Financial Services, American Bankers Association, 1999.

[Eea08] European Environment Agency, 2008.
`http://www.eea.europa.eu`

[Elga85] T. ElGamal, "A public-key cryptosystem and a signature scheme based on discrete logarithms", *IEEE Transactions on Information Theory*, vol. IT-31, no. 4, pp. 469-472, 1985.

[Euro08] European Comission - Information Society, "Internet Communications", 2008. `http://ec.europa.eu`

[Face08] Facebook, 2008. `http://www.facebook.com`

[Fost02] I. Foster, "The Grid: a new infrastructure for 21st century science". *Physics Today*, pp. 42–47, 2002.

[Fouq01] P. A. Fouque and J. Stern, "Fully distributed threshold RSA under standard assumptions", *Lecture Notes in Computer Science*, vol. 2248, pp. 310–330, 2001.

[Gama06] C. Gamage, B. Gras and A.S. Tanenbaum, "An identity-based ring signature scheme with enhanced privacy", *Proceedings of the IEEE SecureComm Conference*, pp. 1–5, 2006.

[Genn96] R. Gennaro, S. Jarecki, H. Krawczyk and T. Rabin, "Robust threshold DSS signatures", *Lecture Notes in Computer Science*, vol. 1070, pp. 354–371, 1996.

[Goll04] P. Golle, D. Greene and J. Staddon, "Detecting and correcting malicious data in VANETs", *Proceedings of the 1st ACM international workshop on Vehicular Ad Hoc Networks*, pp. 29–37, 2004.

[Guo07] J. Guo, J.P. Baugh and S. Wang, "A group signature based secure and privacy-preserving vehicular communication framework", in *Mobile Networking for Vehicular Environments*, pp. 103–108, 2007.

[Jadi04] P. Jadia and A. Mathuria, "Efficient Secure Aggregation in Sensor Networks", *Lecture Notes in Computer Science*, vol. 3296, pp. 40–49, 2004.

[Jhu07] J. Hu, "Trust management in mobile wireless networks: security and survivability", Ph.D. Thesis, The Florida State University, 2007.

[John01] D.J. Johnson, A.J. Menezes, S.A. Vanstone, "The Elliptic Curve Digital Signature Algorithm (ECDSA)", *International Journal of Information Security*, vol. 1, pp. 36-63, 2001.

[Libe05] D. Liben-Nowell, "An algorithmic approach to social networks", Ph.D. Thesis, MIT Computer Science and Artificial Intelligence Laboratory, 2005.

[Lin07] X. Lin, X. Sun, P.-H. Ho and X. Shen, "GSIS: A Secure and Privacy Preserving Protocol for Vehicular Communications", *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6, pp. 3442–3456, 2007.

[Mica08] Crossbow Technology, Inc., "MICA2 Data Sheet", 2008. `http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICA2_Datasheet.pdf`

[Mikr07] A. J. Mikroyannidis, "Towards a social semantic web", *Computer*, vol. 40, no. 11, pp. 113–115, 2007.

[Milg67] S. Milgram, "The small world problem", *Psychology Today*, vol. 2, pp. 60-70, 1967.

[Mill99] C. K. Miller, "Multicast Newtorking and Applications", *Reading MA: Addison Wesley*, 1999.

[Netw08] The Network Simulator - ns, 2008. `http://nsnam.isi.edu/nsnam/index.php/Main_Page`

[News04] J. Newsome, E. Shi, D. Song and A. Perrig, "The Sybil attack in sensor networks: analysis & defenses", *Proceedings of the third international symposium on Information Processing in Sensor Networks*, pp. 259–268, 2004.

[Nico04] A. Nicolosi and D. Mazieres, "Secure acknowledgement of multicast messages in open peer-to-peer networks", *3rd International Workshop on Peer-to-Peer Systems - IPTPS'04*, 2004.

[Okam98] T. Okamoto and S. Uchiyama, "A new public-key cryptosystem as secure as factoring", *Lecture Notes in Computer Science*, vol. 1403, pp. 308–318, 1998.

[Oste07] B. Ostermaier, F. Dötzer and M. Strassberger, "Enhancing the security of local danger warnings in VANETs - A simulative analysis of voting schemes", *Proceedings of the The Second International Conference on Availability, Reliability and Security*, pp. 422–431, 2007.

[Pan07] J. Pan, L. Cai, X. Shen, J. W. Mark, "Identity-based secure collaboration in wireless ad hoc networks", *Computer Networks*, vol. 51, no. 3, pp. 853–865, 2007.

[Parn05] B. Parno and A. Perrig, "Challenges in securing vehicular networks", *Proceedings of the ACM Workshop on Hot Topics in Networks*, 2005.

[Pkcs08] RSA Laboratories, "Public-Key Cryptography Standards (PKCS)", 2008. `http://www.rsasecurity.com/rsalabs/node.asp?id=2124`

[Przy03] B. Przydatek , D. Song, A. Perrig, "SIA: secure information aggregation in sensor networks", *Proc. 1st International conference on Embedded networked sensor systems*, pp. 255–265, 2003.

[Quin01] B. Quinn and K. Almeroth, "IP multicast applications: challenges and solutions", *Internet RFC 3170*, 2001. `http://www.ietf.org`

[Raba07] N. M. Rabadi and S. M. Mahmud, "Performance evaluation of IEEE 802.11a MAC protocol for vehicle intersection collision avoidance system", *Consumer Communications and Networking Conference - CCNC 2007*, pp. 54–58, 2007.

[Rabi79] M.O. Rabin, "Digitalized Signatures and Public-key Functions as Intractable as Factorization", *MIT Technical Report, MIT/LCS/TR-212*, 1979.

[Raja05] H. Rajasekaran, "An incentive based distribution system for DRM protected content using peer-to-peer networks", *1st International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*, pp. 150–156, 2005.

[Raya06a] M. Raya, P. Papadimitratos, and J.-P. Hubaux, "Securing vehicular communications", *IEEE Wireless Communications Magazine*, vol. 13, no. 5, pp. 8–15, 2006.

[Raya06b] M. Raya, A. Aziz and J.-P. Hubaux, "Efficient secure aggregation in VANETs", *Proceedings of the 3rd International Workshop on Vehicular Ad hoc Networks - VANET'06*, pp. 67–75, 2006.

[Raya07a] M. Raya and J.-P. Hubaux, "Securing vehicular *ad hoc* networks", *Journal of Computer Security, Special Issue on Security of Ad Hoc and Sensor Networks*, vol. 15, no. 1, pp. 39–68, 2007.

[Raya07b] M. Raya, P. Papadimitratos, I. Aad, D. Jungels and J.-P. Hubaux, "Eviction of misbehaving and faulty nodes in vehicular networks", *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1557–1568, 2007.

[Rive96] R. L. Rivest and A. Shamir, "PayWord and MicroMint: two simple micropayment schemes", *Lecture Notes in Computer Science*, vol. 1189, pp. 69–87, 1996.

[Rsa78] R. L. Rivest, A. Shamir and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems", *Communications of the ACM*, vol. 21, pp. 120–126, 1978.

[Saba06] J. Sabater-Mir, "Towards the next generation of computational trust and reputation models", *Lecture Notes in Computer Science*, vol. 3885, pp. 19–21, 2006.

[Saha04] A. K. Saha and D. B. Johnson, "Modeling mobility for vehicular *ad hoc* networks", *Proceedings of the 1st International Workshop on Vehicular Ad Hoc Networks-VANET'2004*, pp. 91–92, 2004.

[Sale03] N. Salem, L. Buttyan, J. Hubaux, M. Jakobsson, "A charging and rewarding scheme for packet forwarding in multi-hop cellular networks", *Proceedings of the 4th MobiHoc*, pp. 13–24, 2003.

[Sebe07a] F. Sebé and J. Domingo-Ferrer, "Scalability and security in biased many-to-one communication", *Computer Networks*, vol. 51, no. 1, pp. 1–13, 2007.

[Sham79] A. Shamir, "How to share a secret", *Communications of the ACM*, vol. 22, pp. 612–613, 1979.

[Shou00] V. Shoup, "Practical threshold signatures", *Lecture Notes in Computer Science*, vol. 1807, pp. 207–220, 2000.

[Staa05] S. Staab, P. Domingos, P. Mika, J. Golbeck, L. Ding, T. W. Finin, A. Joshi, A. Nowak and R. R. Vallacher, "Social networks applied", *IEEE Intelligent Systems*, vol. 20, no. 1, pp. 80–93, 2005.

[Stra04] T. Straub and A. Heinemann, "An Anonymous Bonus Point System For Mobile Commerce Based On Word-Of-Mouth Recommendation",

Proceedings of the 2004 ACM Symposium on Applied Computing, pp. 766–773, 2004.

[Vass03] J. Vassileva, "Motivating Participation in Peer to Peer Communities", *Lecture Notes in Computer Science*, vol. 2577, pp. 141–155, 2003.

[Vide08] Videntity, 2008. `http://videntity.org`.

[Vish03] V. Vishnumurthy, S. Chandrakumar, and E. Sirer, "Karma: A secure economic framework for p2p resource sharing", *Workshop on Economics of Peer-to-Peer Systems*, 2003.

[Wang06] D.-W. Wang, C.-J. Liau and T. Sheng Hsu, "Privacy protection in social network data disclosure based on granular computing", *IEEE International Conference on Fuzzy Systems*, IEEE Computer Society, pp. 997–1003, 2006.

[Weim04] A. Weimerskirch and D. Westhoff, "Zero common-knowledge authentication for pervasive networks", *Lecture Notes in Computer Science*, vol. 3006, pp. 73–87, 2004.

[Wolf03] T. Wolf and S.Y. Choi, "Aggregated hierarchical multicast - a many-to-many communication paradigm using programmable networks", *IEEE Transactions on Systems, Man and Cybernetics - Part C*, vol. 33, no. 3, pp. 358–369, 2003.

Alexandre Viejo Galicia

June 2008