CONTRIBUTIONS TO PRIVACY PROTECTION FOR UBIQUITOUS COMPUTING

Pablo A. Pérez-Martínez

Dipòsit Legal: T 64-2015

DOCTORAL THESIS

Pablo A. Pérez-Martínez

Contributions to Privacy Protection for Ubiquitous Computing

Universitat Rovira i Virgili

Pablo A. Pérez-Martínez

Contributions to Privacy Protection for Ubiquitous Computing

DOCTORAL THESIS

Supervised by Dr. Agusti Solanas

Department of Computer Engineering and Mathematics

Universitat Rovira i Virgili

Tarragona
2015

UNIVERSITAT ROVIRA I VIRGILI

FAIG CONSTAR que aquest treball, titulat " Contributions to Privacy Protection for Ubiquitous Computing ", que presenta Pablo Alejandro Pérez Martínez per a l'obtenció del títol de Doctor, ha estat realitzat sota la meva direcció al Departament d'Enginyeria Informàtica i Matemàtiques d'aquesta universitat.

HAGO CONSTAR que el presente trabajo, titulado " Contributions to Privacy Protection for Ubiquitous Computing.", que presenta Pablo Alejandro Pérez Martínez para la obtención del título de Doctor, ha sido realizado bajo mi dirección en el Departamento de Ingeniería Informática y Matemáticas de esta universidad.

I STATE that the present study, entitled " Contributions to Privacy Protection for Ubiquitous Computing ", presented by Pablo Alejandro Pérez Martínez for the award of the degree of Doctor, has been carried out under my supervision at the Department of Computer Engineering and Mathematics of this university.

Agustí Solanas

Tarragona, 25 d'Agost de 2015

El/s director/s de la tesi doctoral
El/los director/es de la tesis doctoral
Doctoral Thesis Supervisor/s

Agustí Solanas Gómez

Abstract

The introduction of the first computers together with the emergence of the Internet was the trigger for a technological revolution that is still evolving. This revolution began with households gaining access to personal computers, capable of communicating with other computers anywhere in the world. As a result, web platforms grew very rapidly reaching homes worldwide and providing information to its users through a network.

Furthermore, continual efforts to increase the computing power while reducing the physical size of microprocessors allowed them to be produced at lower costs, and aided the advancement of digital communications as they could be built into almost any device or object used in our everyday lives. This is how the concept of Ubiquitous Computing was born (also called Pervasive Computing), whereby the objects around us can have an embedded computer and establish communications with each other, in order to provide personalized services to assist with our tasks, while at the same time they are connected to the Internet and so can share that data with other devices around the world in real time.

Since this has become possible, scientists and technology companies worldwide have focused their efforts on developing this concept further, using the computer to understand speech, recognize facial expressions and interact with humans in an intelligent way, making our lives easier, particularly for the elderly and disabled.

However, because it is possible to have computers almost anywhere and within any object, environments become active, and this has opened up new discussions on issues such as privacy and security, particularly when it comes to devices capable of self-localization where the location of their users is known all the time. An Active Environment can be any space with this technology (e.g. a shop, a room, a kitchen, etc.) that has the ability to interact with the user. This technology is expanding so rapidly. Smart cities are starting to emerge, because the entire city is connected and therefore is considered as an active environment. When we combine Ubiquitous Computing with Active Environments we get the basis of what is now known as Ambient Intelligence. This dissertation focuses on the privacy and security issues introduced by Active Environments and Ubiquitous Computing from a global perspective.

Privacy has been the subject of numerous publications in recent years, considered from many different views, such as the legal, social, economic and technological development perspectives, all taking an increasingly significant importance in today's world. Many devices like smartphones, tablets, smart watches, and even TVs, store information about the users and their activities. Misuse of this data, whether intentional or unintentional, has led to scandals worldwide, such as big companies disclosing personal data by mistake; big brands obtaining the location of users against their will; and governments accessing the data of unknowing users' personal computers and phones. This abundance of data and its potential misuse makes the protection of privacy a very important issue in the development of future technologies.

In this dissertation, we firstly present some of the main privacy issues in mobile systems, such as Location Based Services (LBS), and we introduce the concept of W^3-Privacy, the three dimensions of user privacy, and how we can achieve this with a simple protocol. We propose four protocols to protect users' privacy, thereby preserving the proper functioning of the service. And we propose a system based on collaboration among users to achieve privacy in LBS, with no need to rely on other users.

Finally, we expose the dangers of some of the current tracking systems for people suffering from mild cognitive impairments, which manipulate very sensitive data (i.e. health status or diseases), and we propose a private system based on role access to the information. After that, we show some of the privacy issues related to Smart Cities, and we introduce the five dimensions model of privacy for citizens, built applying some of the current private models.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Dr. Agusti Solanas for believing in me, for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I will never forget our long days in Rome and Padova working so hard to make it possible.

My sincere thanks also go to Dr. Roberto di Pietro and Dr. Mauro Conti, for offering me opportunities in their research group and guiding me working on diverse exciting projects.

I am very grateful to my other collaborator and friend, Toni Martínez-Ballesté, joint work with whom appears in this thesis. I have learned so much from working together with you, and am grateful for the ideas and energy you have given me.

I thank my classmate and friend Albert Fernández Mir, for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had during our studies. Also I thank my friend Iñigo Goiri, employee at Microsoft company, for his help during the first years of career.

I really appreciate the effort of Teresa, Johanna, James, Ellie and Samuel for their support to correct this dissertation few days before submitting it. I am glad to work with people like them.

Last but not least, I would like to thank my family: my parents Antonio and Ana, for believing in me, even when nobody else did it, for supporting me throughout my life and for their huge efforts to give me all I needed, my sister and brother Anabel and Antonio, for making me proud to have both of you as siblings, and my girlfriend Sara, for her huge support and encouragement, and for not letting me give up when I was tired, I could not have done it without her support.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

This chapter introduces the issues we are facing in this dissertation. In addition, it briefly describes the solutions we propose to tackle those issues. Finally, the structure and organisation of the present thesis are outlined.

Contents

1.1 Motivation

Ubiquitous computing has received both academic and commercial interest. In [24] a detailed description of the challenges for ubiquitous computing is presented from different perspectives, but we mainly focus on privacy concerns.

Definition 1. *A location-based service (LBS) is a software application for a IP-capable mobile device that requires knowledge about where the mobile device is located.*

Due to the improvements and the quick evolution of the technology, more and more systems make use of the location, thanks to the capabilities of self-localisation, by means of Global Positioning System (GPS), Global Navigation Satellite Systems (GLONASS) or Galileo (the European positioning satellite). For instance, thirty years ago the first mobile phone appeared, and in few years they have been evolved until they became devices with almost the same capabilities and computational power as a personal computer. This has meant that the location-based services have grown exponentially in recent years.

Location-based services offer many opportunities for the mobile users, as for example:

- Requesting the nearer service or business, such as a pharmacy or a restaurant.
- Receiving alerts, such as notification of sale on gas or information about the transit.
- Finding friends or people with cognitive impairments.

LBS providers need the private information of the users in order to give the best service possible. In the case of a service that shows the weather information, the location sent to the provider could be just a city, or a big area where the user is located and it would be enough to get the service, but if the user needs to know where is the nearest pharmacy, he should share his exact location in order to get a proper result. Even more so, in case of a person suffering from mild cognitive impairments (MCI), the system constantly needs the exact location of the user, the heart rate, blood glucose level, etc., because the life of the user could depend on the proper functioning of the service.

But this raises questions related to privacy. Why should we share our private data? Can we trust all the service providers that ask for our data? What if the service provider makes a dishonest use of our data? Or, even if the it is honest, what if an attacker gets these data? The private data of users should be private even for the service provider to prevent a dishonest use, specially in cases where the information about the health of the user is shared.

These privacy concerns not only affect to mobile devices. New radio identification systems such as RFIDs which are able to identify objects with no need of contact.

Definition 2. *Radio frequency identification (RFID) is a technology that uses electronic tags placed on objects, people, or animals to relay identifying information to an electronic reader by means of radio waves [65].*

RFID is aimed at efficiently identifying products without the need for visual contact that has greatly influenced the manufacturing businesses in recent years. This technology has been called to replace bar codes and to improve the management of products. Although the RFID technology has been widely accepted by the manufacturing and retailing sectors, there are still many issues regarding its scalability, security and privacy. Also related to privacy, the sharing of identification information amongst multiple parties is an issue (specially after the massive outsourcing that is taking place in our global market). Securely and efficiently sharing identification information with multiple parties is a tough problem that must be considered so as to avert the undesired disclosure of confidential information.

As the authors described in [62], RFIDs have two main privacy concerns for users: *clandestine tracking* and *inventorying*. RFID tags react to reader interrogation without the owners' knowledge. Thus, if a reader is close enough, clandestine scanning of tags is a realistic threat. Most RFID tags reply with unique identifiers, even tags that protect data with cryptographic algorithms. In consequence, somebody carrying an RFID tag broadcasts its serial number to nearby readers, allowing for clandestine physical tracking.

When the identifier of a tag is combined with some kind of personal information, the threat to privacy grows. For example, when a consumer makes a purchase with a credit card, the shop can link his identity with the serial number of the credit card's tag. Marketers can then identify and profile the consumer using networks of RFID readers.

In addition to their identifiers, some tags carry information about the items to which they are attached. These tags include a field for the "General Manager", typically the manufacturer of the object, and an "object class", typically a product code, known formally as a stock-keeping unit (SKU). Thus, somebody carrying an RFID tag is subject to clandestine inventorying. A reader can silently determine what objects a person is carrying, and show personal information like the types of medications he is carrying and, therefore, what illnesses he may suffer from; the RFID-enabled loyalty cards he carries and, therefore, where he shops; his clothing sizes and accessory preferences, etc. This problem of inventorying is largely particular to RFID.

This technology is already adopted in several scenarios; we show some of the most common uses:

- *Toll-payment transponders:* Automated toll payment transponders consist on small plaques positioned in windshield corners, this technology is commonplace worldwide.

- *Libraries:* Some libraries have implemented RFID systems to facilitate book checkout and inventory control and to reduce repetitive stress injuries in librarians.

- *Passports and IDs:* An international organization known as the International Civil Aviation Organization (ICAO) has promulgated guidelines for RFID-enabled passports and other identifier documents.

- *Human implantation:* Few other RFID systems have inflamed the passions of privacy advocates like the VeriChip system [3]. VeriChip is a human-implantable RFID tag, much like the variety for house pets. One intended application is medical-record indexing; by scanning a patient's tag, a hospital can locate her medical record.

- *Credit, public transportation, and some other cards:* Most of the current cards carry an RFID tag in order to pay remotely, in a shop, bus, metro or wherever they have implanted the system.

In the context of mobile-phones, it is becoming more common to have a smartphone, and it is pretty hard to find somebody who does not have one. The use of smartphones also allows people to connect a body sensor in order to store and analyse their vital signs and track their improvements. Electronic health (e-health) and more recently mobile health (m-health) are focussing the attention of the public and private healthcare sectors. Their main aim is to help reducing economic costs whilst maintaining or even increasing the assistance quality that patients receive.

Body sensors are the linchpin of e-health and m-health since they play a fundamental role in gathering biomedical information that is essential for the proper remote supervision of patients. Notwithstanding, body sensors can no longer be understood as individual sensors that locally collect data, but as cooperative sensors organised in networks (BSN) that share biomedical information in complex infrastructures based on advanced information and communication technologies (ICT).

The new paradigms of cloud storage and cloud computing open the door to the migration of body sensor networks from local infrastructures to more powerful and globally accessible cloud services. However, due to the fact that the data collected by BSN are very sensitive, cryptographic mechanisms should be provided in order to assure the protection of patients privacy.

Moreover, most of these systems are not capable of self located, but RFID readers can be deployed in a restricted area, so each reader is located in a determined location, and it can identify RFID tags in order to know their location. If a set of RFID readers are deployed in a city, the system could track the citizens and perform a profile of each of them, in order to use this information for spam, to sell it to use it in order to deny an insurance to somebody, etc., and this could suppose a huge risk for the privacy of the users. A city where hundreds or thousands of RFID readers are deployed and citizens use RFID tags for different services could seem a futuristic situation, but in the last years the concept of "Smart City" is being expanded among the biggest cities of the world, with the aim of improving the quality of life of its citizens. In these cities, they already have enough RFID readers deployed in order to offer services to their citizens.

Countries are making great efforts to be competitive, attract investments and talent, reduce debt and be more sustainable. Both countries and cities struggle for competitiveness, which are competing at an international level for investments, talent and quality of life, and they realise that the most promising path to success is the use of technology. Specifically, information and communication technologies (ICT) allow local governments and companies to develop ubiquitous innovative solutions that improve city operations in a variety of areas, such as transportation, energy, sustainability, e-governance, economy and communications.

In big cities, factors related to economies of scale help to reduce operational costs. However, managing big cities is challenging because the number of inhabitants grows steadily and the infrastructures and operational procedures have to be adapted to a growing and very demanding population.

In this context, local administrations have the need for smart procedures to improve the quality of life and the management of resources in cities. As a result of these needs, the concept of *Smart City* was adopted, but in many cases the meaning given to this term changes from person to person. Moreover, the term has gained a kind of marketing value that local governments want to benefit from. Thus, the definition of the term is frequently modified so as to adapt to the needs of the people using it in a particular situation. As a consequence, a number of different definitions and conceptual ideas regarding *Smart Cities* can be found in the literature.

Furthermore, the concerns are even bigger when we are talking about the welfare of people. Age-related diseases are becoming more prominent due to life expectancy increase in developed countries. Mild cognitive impairment and several types of dementia like Alzheimer's disease are gaining importance both socially and economically. Patients suffering from these diseases have different degrees of autonomy and, thus, different needs. Often, relatives or friends take care of those patients. However, during the first stages of the disease, they still have a high degree of autonomy and frown on the supervision of others. Despite their autonomy, patients could get lost and disoriented. Rapidly determining the location of a lost patient is paramount to reduce the risk of suffering serious injuries. Current solutions to this problem are based on the continuous monitoring of the patient. Such continuous control might be seen by most people as a privacy invasion, and it may discourage patients from using these solutions.

1.2 Contribution

Location-based services are becoming more and more popular among the mobile app developers. Most people have a smartphone with self-location and data connection. These capabilities allow users to make use of LBS in order to find a nearby service or place, interact with other users or simply check information. This technology can be very useful, but it entails several privacy risks.

Also the RFID technology can help to improve our processes by simplifying paperwork, speeding up the cataloguing of products and reducing costs. However, the massive use of this technology implies the management of billions of tags (which is a challenge) and might suppose a thread in terms of security and, also, in terms of privacy.

In this dissertation, we concentrate on these issues (*i.e.* Security, privacy and scalability in LBS, RFID, sensor networks and smart cities). Specifically, our contributions are the following:

- We organise, classify, and analyse the main privacy protection protocols in location-based services.
- We propose different systems able to guarantee user's privacy. In one hand we present a couple of them based on collaboration, where the users collaborate among them to achieve privacy. In the other hand, we also propose two more non-collaborative protocols.
- We introduce the concept of W^3-Privacy, which defines the three dimensions of user privacy (Where, What, Who).
- We propose an intelligent and private public parking system for smart cities.
- With the aim to avert privacy and security risks and simultaneously gain the full advantages of body sensors networks (BSN) in the cloud, we propose a complete cryptographic framework that grants users/patients full control over their data whilst, at the same time, allowing the controlled sharing of information. Our solution is based on the concept of Privacy as a Product (PaaP) and the Raikova-Vo-Bellovin-Malkin (RVBM) scheme.

- We recognise the growing importance of body sensor networks and we show that their use is paramount for the proper evolution of our modern society towards a better and more accessible health system. However, BSN can no longer be seen as isolated systems that send information to authorised parties only. On the contrary, they are sharing their collected data with cloud services so as to take advantage of cloud storage and computational benefits. As a result, if the proper cryptographic measures are not taken, BSN connected to the cloud could represent a privacy and security threat for their users.

Finally, due to the proliferation of the concept of "Smart City", we define the *Citizens Privacy* concept, which defines the five dimensions model for smart cities.

1.3 Organisation

Chapter 1 introduces the concept of privacy in ubiquitous computing, starting with the motivation that has led to this dissertation and the contributions to avoid the privacy concerns. The rest of this dissertation is organised as follows:

Chapter 2 shows the current privacy protection techniques in location-based systems, RFIDs, body sensors and smart cities. Then, chapter 3 and 4 present protocols to protect privacy in LBS through the collaboration of users.

Concluding with location-based services. In chapter 5 we present the concept $W3$-Privacy, which includes the three dimensions of privacy in LBS to be met to ensure privacy.

In the second part, in chapter 6, we apply a scalable method for RFID identification to a smart city, allowing the smart and private use of public parkings. Chapter 7 presents a framework that enables private sharing of information obtained by body sensors, allowing users to share and sell the information.

Chapter 8 introduces the concept "Citizen Privacy" which includes the 5 dimensions of privacy that must be met in a smart city to guarantee the privacy of its citizens and chapter 9 presents the "mobile carer", an intelligent, private and autonomous system for mobile phones able to help a person with mild cognitive impairments (MIC) in case of emergency. Finally, this thesis ends with conclusions and the most important contributions.

Chapters on Parts I, II and III start with a brief summary of the chapter's content, then an introduction to the current chapter followed by the proposed protocol and finally some results and conclusions.

Chapter 2
Background and Related Work

This chapter starts by describing the current state-of-the-art privacy protection in Location Based Services. We organise, classify and analyse the most common protocols. We then present the RFID technology and some of the main privacy-preserving techniques. The chapter finally shows the concept of Privacy-as-a-product and the most relevant contributions of privacy-preserving in the Cloud and body sensors.

Contents

2.1 Location Based Services

In the simplest form of communication between an LBS user and an LBS provider (Figure 2.1), the former sends her location (x, y) to the latter along with an identifier (ID) and a request for information (e.g. [x, y, ID, “Where is the closest bus station?”]). Then, the provider answers the user with the required information.

If the data is sent as plain text, an attacker listening to the communication channel would get all the transmitted data, but lets suppose that the server provider adds some minor protection techniques to the service and works under an encrypted channel, so an attacker, by means of a man in the middle attack, would not see the data. Even with encrypted communication, the system can not guarantee the privacy of the user from the provider because the latter knows the real location of the former. Thus, in this scheme users must trust providers.

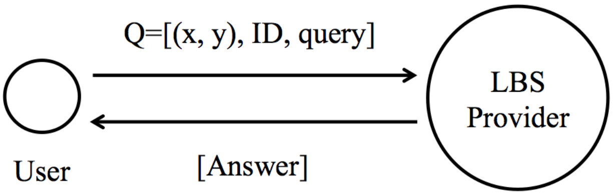

Figure 2.1: Simplest form of communication between user and service provider.

To give an example, imagine a user using Google Maps [52] from his smart phone. The user is looking for a pharmacy and asks the application for the nearest one. Google Maps needs the location of the former, so this user sends private data to Google [51], which will use the data to profile the user and profit. Then, when the same user accesses Google web search, spam about medicines, health care centres, pharmacies, among others, can appear.

In an attempt to protect users' privacy in location-based services, policies that would defend the rights of users were created, although they do not prevent third parties from accessing the private data that users do no want to share. Thereby, research on this topic has motivated the creation of a huge number of protocols to confront this problem.

The most obvious system to face this problem is to replace the identifier of users with a pseudonym, but this assumes that the same user always has the same identifier and can be thus identified. In an attempt to solve this problem, Beresford and Stajano [14] propose that the pseudonym should change in order to prevent the server from linking pseudonyms with the real identities of users, but Bettini *et al.* [15] show that with a sequence of locations they can be linked quite easily.

The most extended privacy system is *Spatial Cloaking*, in which the user's location is replaced by geographical areas that can contain more users, but there are also systems based on cryptography and obfuscation of locations.

Most of the *Spatial Cloaking* methods are based on a Trusted Third Party (TTP), [41], [42], [15] which is responsible for managing the private data of users in order to create an area containing a set of users. This area can be used by the user in order to submit queries to the provider, so the provider does not know the real location of the user within the area. Another option is to allow the TTP to submit the queries and then send the answers to the users, but these kinds of protocols also have the typical drawbacks of a centralised system (bottlenecks, single points of attack, etc.), so they can jeopardize the integrity of the system and the users' privacy. Moreover, these protocols do not achieve privacy protection since users must trust the TTP and share their private data with it.

There are different proposals to compute the area in *Spatial Cloaking* methods, some of which are based on the *k-anonymity* property [53], a concept introduced by

Gruteser and Grunwald [53] and defined as following:

Definition 1. *Let $RT(A_1, \ldots, A_n)$ be a table and QI_{RT} be the quasi-identifier associated with it. RT is said to satisfy k-anonymity if and only if each sequence of values in $RT["QI_{RT}]$ appears with at least k occurrences in $RT[QI_{RT}]$.*

On the other hand, there are proposals based on *obfuscation* [36] (see Definition 2), in which the degradation of users' location reduces the risks of privacy loss.

Definition 2. *Obfuscation is the process of degrading the quality of information about a person's location, with the aim of protecting that person's location privacy.*

Duckham and Kulik [36] introduced the concept with two parameters, *innacuracy*, which implies giving a fake location instead of a real one, and *imprecision*, which implies giving plurality of possible locations. In [53] the authors added to the idea of temporal obfuscation the concept of *temporal cloaking*, where a user can specify the parameters of delay, although this protocol forces the users of the same area to have the same requirements of privacy. In [42], *Clique Cloak* allows the users of a same area to have different requirements of privacy, but the algorithm is not scalable for big groups.

Since centralised systems are vulnerable, new systems with no need for TTP have been proposed. Some of them make use of the collaboration between users through ad-hoc networks [109], [47], [50], [16]. In these systems, users communicate amongst them in order to produce obfuscation areas with different algorithms. In [109], users use privacy homomorphisms in order to hide their locations but are able to create areas including all participating users. In [93], a decentralised system is proposed where users collaborate with no need for trusting other users, since private data is hidden from collaborative users, thanks to public key cryptography. In this system, users perform a chain of h users (number of hops of the message) through the Internet, whereby the message hops until it reaches the last user. The last user is then responsible for submitting the query to the provider. Because no users know the owner of the query, the provider will not link the query with an identity, even in the case of collusion between users and the provider.

Other kinds of proposals that do not need either TTP or collaboration of users are PIR protocols (Private Information Retrieval). These protocols are very complicated systems studied by the database and cryptography communities, which require an encrypted database. The aim of PIR is to allow a user to get the record i from a database without revealing i. In [48, 67], the authors use this concept for LBS, yet the biggest drawback of these protocols is the huge computational complexity and the overhead that they add in comparison with traditional methods.

As we stated before, a huge number of methods have been proposed. For the sake of clarity, we have classified those methods depending on the way they manage the location of users. First, we consider the methods that do not distort locations, and then we move to the ones that use approximations. See Figure 2.2 for a graphical representation of this classification.

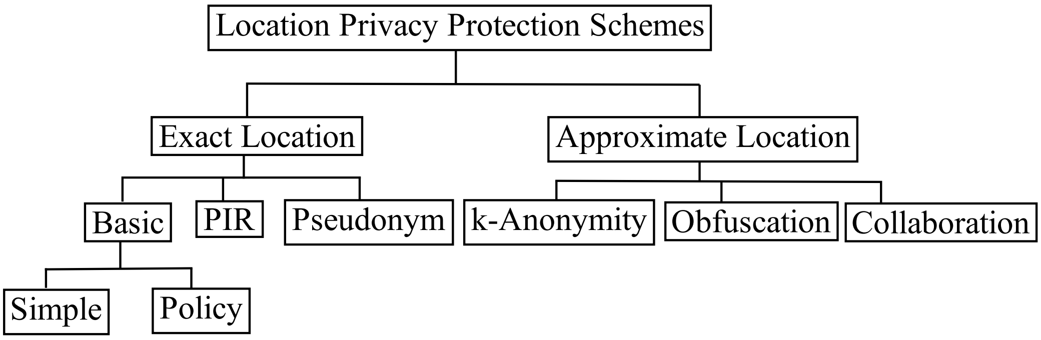

Figure 2.2: Classification of location privacy methods according to the way they manage locations [93].

2.1.1 Exact location schemes

We classify all the methods that do not distort the location of LBS users to protect their privacy under exact location schemes category. These methods are very common because most of them are conceptually simple. In addition, as they do not modify the location of users, the obtained results are optimal. Apart from the basic/simple scheme described above, we consider three non-disruptive schemes: (i) Policy-based schemes; (ii) Pseudonymizers; and (iii) Private Information Retrieval (PIR)-based schemes.

Policy-based schemes: The architecture of these schemes is like the simple scheme with a single user and a provider.

In this scheme (Figure 2.3), the communication protocol is not modified, so the communication between the user and the provider is exactly the same as in a simple communication. However, in this case the provider adheres to a set of privacy policies [104]. Consequently, the user has the right to ask for a compensation if the provider does not fulfill his duties. The Geopriv (Geographic Location/Privacy) Charter of the IETF supports the use of policies [101].

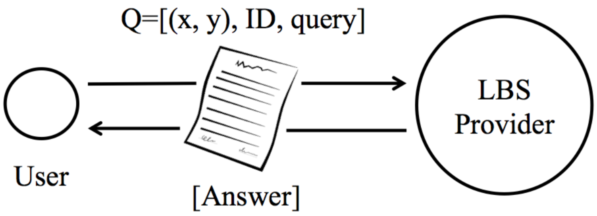

Figure 2.3: Policy-based scheme, where the provider adheres to a set of privacy policies.

In an environment where a number of policies need to coexist, several policies can be in conflict, either because of a specification error or because of application-specific constraints. It is therefore important to provide the means of detecting conflicts in the policy specification [71, 79]. According to [55], by considering the different conflict types, it is possible to define rules that can be used to recognise conflicting situations in the policy specification. We can find some examples in the current literature about how these rules can be used as part of a detection process in [25, 26]. Another issue regarding policy-based systems is whether the policies should apply to all users and how their preferences are respected. In [19], the authors consider cases where no absolute control from an authority is accepted, while in [18], a "promise theory" attempts to provide "political autonomy" to entities and decentralize policy management.

Policy-based schemes are widespread among the main LBS applications, and offer some protection to the user's privacy against the service provider, but it does not prevent her against an external attack, or even the provider, which can make a mistake and share the private data of its users. For instance, in 2006 AOL [1] released detailed search logs of a large number of its users. The release was intentional and intended for research purposes; however, the public release implied that all the Internet users could see the results, rather than a select number of academics. AOL did not redact any information, thus causing privacy concerns since users could be potentially identified by their searches.

Pseudonymizers: These schemes add a trusted third party (a pseudonymizer) to the basic model. The pseudonymizer, shown in Figure 2.4, mediates between users and providers. Users send their queries to the pseudonymizer, which replaces the real identity of the users (e.g. their IP addresses) with a pseudonym. In this way, providers cannot identify users because they become hidden behind the pseudonymizer.

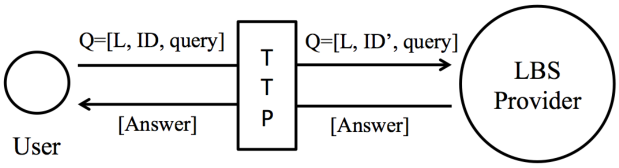

Figure 2.4: Pseudonymiser mediates between users and providers. Users send the queries to the pseudonymiser, which is responsible for submiting these queries to the server provider.

In [99], the authors suggest user-generated random pseudonyms. The user provides his location data together with a timestamp and the associated pseudonym to the location service. The use of pseudonyms as identifiers allows for anonymity in identity management. An essential factor for effectiveness of pseudonyms is the unlinkability between the pseudonym and its holder and whether pseudonyms can be linked between each other. Notwithstanding, users *must trust* pseudonymizers because *they have full access to their real locations and identities.* Also, if users send several queries from the same location (e.g. from their residence), providers can determine their real identities by using e.g. a public telephone directory or just observing who is in that location. These attacks are known as Restricted Space Identification (RSI) and Observation Identification (OI) [53].

PIR-based schemes: Private information retrieval (PIR) is a difficult problem mainly studied by the database and cryptography communities. The goal of PIR is to allow a user to obtain a record (i) from a database without revealing i.

The PIR problem was first formulated and studied in [27], where the solutions assumed multiple databases and were aimed at information-theoretical security. However, the assumption that multiple databases would not communicate with one another is considered unrealistic in practical applications. Later in [28, 46, 120], PIR schemes with a single database were proposed. These solutions were based on computational complexity assumptions, such as the difficulty of factoring $n = pq$. Unfortunately, the computational costs of these solutions are very large due to their bit-by-bit processing nature. They require O(N) multiplications (mod. a 1024-bit number) for retrieving just one bit, where N is the total number of bits in the database [12].

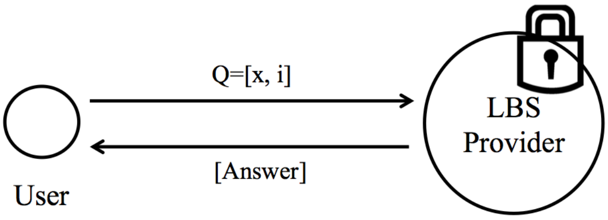

Figure 2.5: PIR-based scheme where the provider has an encrypted database.

Most PIR schemes (Figure 2.5) based on the computational complexity assumption aim at reducing the cost of communications. The scheme in [21] can achieve a communication cost of polylog(N) while those cited previously have exponential communications. Mathematically, these schemes are very beautiful. But from the implementation point of view, they are completely impractical since they all have computation complexity of at least O(N) to retrieve just 1 information bit. A practical scheme should process messages file-by-file instead of bit-by-bit.

In [49, 67], the authors use the concept of PIR in LBS. The main problem of these methods is their high computational complexity. In addition, the LBS provider must implement very sophisticated protocols to exchange information with users. Nowadays, these proposals seem impractical.

2.1.2 Approximate location schemes

The methods that distort the real locations of users assume that the modification of the locations prevents the provider from learning the private information of users, sacrificing accuracy in the response. We consider three main categories: (i) k-Anonymizers; (ii) Obfuscation; and (iii) Obfuscation by collaboration (Collaboration for short).

k-Anonymizers: As we can see in Figure 2.6, there are TTPs to which users send their queries. After collecting some queries, k-anonymizers build groups of k users and compute a fake location (e.g. a centroid) that represents all the members of the same group. Then, the real locations are replaced by the centroid of the group and the provider cannot distinguish which user in the group sent the query. Although the k-anonymity property is very interesting and increases the privacy level of users, this approach has all the problems of the TTP-based approaches and, in addition, the obtained results are not accurate.

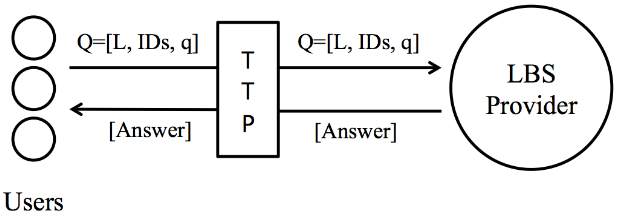

Figure 2.6: k-Anonymisers are able to compute a region where k are located.

According to [14], anonymity can be seen as “a state of being not identifiable within a set of subjects, the anonymity set”. Initially, k-anonymity was proposed by Samarati in [100] to protect microdata. The main idea of k-anonymity applied to LBS is to hide a user amongst $k-1$ other users.

In [41, 53], the authors adopt the k-anonymity model and proposed a quad-tree based cloaking algorithm. They assume a static anonymity requirement k_{min} for all users. To achieve k-anonymity, the algorithm recursively subdivides the area around a user's location into four quadrants until the number of users in the area falls below k_{min}, and then returns the previous quadrant as the cloaking region. This technique does not differentiate the privacy requirements of different users. Moreover, no restriction is imposed on the cloaking region size. Thus, a cloaking region can be very large, which may lead to an inaccurate query result and poor service quality.

Gedik and Liu [42] proposed the technique of supporting personalized privacy requirements, thus capturing the privacy and QoS requirements on a per-user basis. The authors presented a cloaking algorithm called *Clique Cloak*, which constructs an undirected graph for all the requests that have not been anonymised yet. Every time the server receives a new request, it attempts to identify a clique involving the new request and some existing requests, and cloak them together with the same region. However, this method has several drawbacks:

- the effectiveness of this method is limited to users with small k values,
- the cost of searching a clique in a graph is high, and
- some requests that cannot be anonymised will be dropped when their lifetimes expire.

These drawbacks affect the user's experience of the service.

There are other proposals in the literature, as for example Casper, proposed in [81]. Casper employed a grid-based pyramid structure to index all user locations.

Besides the anonymity level k, a user can specify A_{min}, indicating that the user wants to hide his location information within an area of at least A_{min}. But this model presents some concerns, such as the similar functionality of k and A_{min} (the higher the value of k is, the larger the cloaking area is). And the cloaking region may expand to be arbitrarily large if k is set to a large value and few users are present nearby. To address this problem, Casper uses a privacy-aware query processor to return a list of candidate query results to the anonymizing proxy, which has to locally refine the actual result from the candidate list. This approach incurs a high query processing cost, a high communication cost, and a high local computation cost.

Obfuscation-based schemes: As the authors described in [9], location obfuscation is complementary to anonymity. In particular, rather than anonymizing user's identities, obfuscation-based solutions assume the identification of users and introduce perturbations into collected locations to decrease their accuracy.

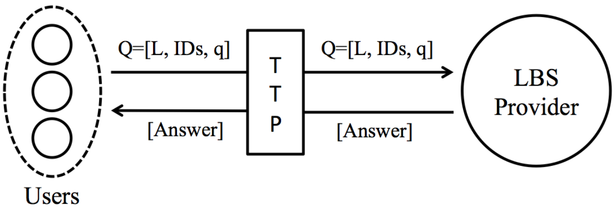

Figure 2.7: Obfuscation-based scheme, where the TTP adds noise to the user's location.

These methods are generally run by a single user and no TTPs are required, and the main idea behind them is to reduce the accuracy of the location. For example, instead of sending the real location, users send a squared area. By doing so, providers just know that a given user is located inside that area but they do not know exactly where. By means of increasing the size of the area, location privacy is also increased but results become worse [9,121]. In Figure 2.7, we can see a graphical example of how TTP collects the location data of users and computes an obfuscated area, where all the users are located.

Current obfuscation-based solutions have some shortcomings:

- They do not provide a quantitative estimation of the actual privacy level, which makes them highly dependent on application contexts and difficult to integrate into a full fledged location-based application scenario.
- Just a single obfuscation technique is usually implemented.

Collaboration-based schemes: In these kinds of methods, the goal is the same as in obfuscation methods and k-anonymizers, that is, to obfuscate the location of a user among other users. However, as we see in Figure 3.4, the strategy is different. Users collaborate to exchange location information that they use to disguise their real location. By collaborating, users avoid TTPs and improve the results of single-user obfuscation methods [29, 105].

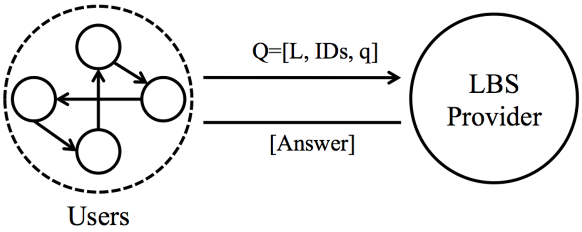

Figure 2.8: In collaboration-based schemes, users collaborate among them in order to submit a query to the service provider.

In [105], the authors present a new free-TTP scheme for privacy protection in LBS. Here, users share their location distorted with Gaussian noise in order to compute a centroid, which will be used to send queries to the service provider.

2.2 Radio Frequency Identification

Radio-Frequency IDentification (RFID) devices have an important presence in our daily life. Even when we do not know it, there are RFID tags in our clothes, mobile devices, etc., and they will become ubiquitous in the near future. The spectacular market push of RFID technology is due to the interest by large retailers, important manufacturers and governments. As a result, almost every object is liable to carry an RFID tag. RFID devices can be seen as a proper substitute to codes that need to be visualized, like bar codes or QR, since they are mainly used to identify objects. Unlike bar codes, RFID devices allow objects to be identified across larger distances, with no need for visual contact, and they help to improve and automate many processes. This is possible due to the ability of RFID tags to be read remotely, fast and in parallel.

2.2.1 Basic scheme of RFID systems

Regardless of their operational frequency, materials or embodiments, RFID systems consist of three main components, namely tags, readers and back-ends:

- **RFID tags** are small devices that can take a variety of possible shapes and embodiments (from stickers to small grains embedded in documents). The most basic RFID tags consist of a microchip and a metal coil. The microchip stores information, and is able to compute some simple operations; the metal coil acts as an antenna that receives information from and sends information to readers. Optionally, RFID tags can carry batteries. In this case they are called *active tags*. Otherwise, they are called *passive tags*. Passive tags are far more common than active tags because they are cheaper (*e.g.* a passive tag costs about $0.05, whilst an active tag might cost about $50 or more). Due to the fact that passive tags do not carry batteries, they harvest energy from

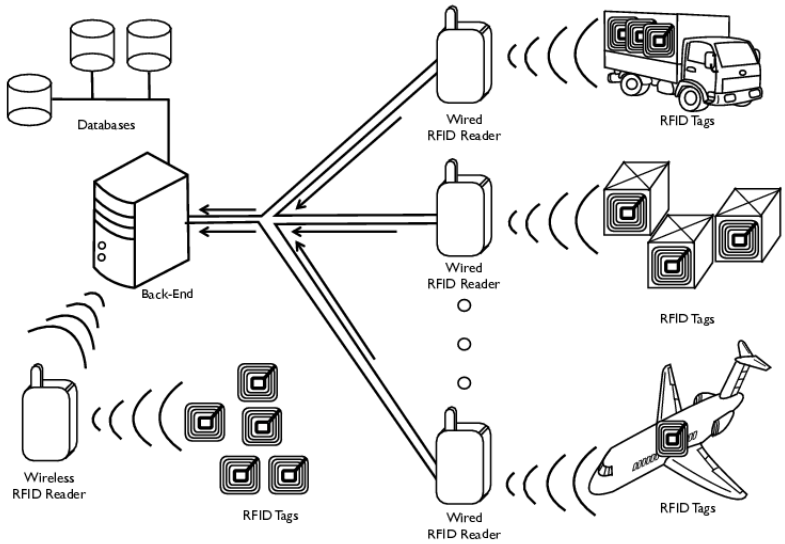

Figure 2.9: Basic components of an RFID system. From left to right: a back-end, RFID readers, and RFID tags. The back-end uses databases to store identification information. RFID readers are used to query RFID tags (which can take a variety of embodiments), retrieve their information, and forward it to the back-end through a wireless or wired channel. Note that in this simplest scheme, RFID readers are used as relays and are not connected amongst themselves [112].

the signal they receive from readers and, consequently, they have very limited storage and computational power.

- **RFID readers** are devices utilized to retrieve information stored in RFID tags. In their simplest operation, readers emit a radio wave so that all tags in their cover range can power up and answer by broadcasting their embedded information (*i.e.* a set of bits[1]). After collecting the information, the readers forward it to a centralized computer (or *back-end*) along with their identification number and a timestamp.

- **Back-ends** are a set of databases connected to computers that receive, decrypt (if necessary), and manage the information collected by RFID readers about RFID tags. Back-ends store all the information required to identify RFID tags. Also, they can have extra information about the products/items to which tags are attached.

In [62], we can find the following classification of RFID tags, which is based on range distances:

[1] In general, these bits represent the electronic product code (EPC) of the item to which RFID tag is attached.

- *Nominal read range:* This is the range indicated by RFID standards and product specifications. It is the maximum distance from which a tag can be read by a reader.
- *Rogue scanning range:* This is the maximum range, beyond legal limits, at which a reader can power and read a tag.
- *Tag-to-reader eavesdropping range:* When a reader powers a tag, a more sensitive receiver can eavesdrop the emissions of a tag without emitting any signal. Thus, the eavesdropping range can be equal to or higher than the rogue scanning range.
- *Reader-to-tag eavesdropping range:* This range is even higher than the previous one because the signal power of the reader is greater than that of the tag.
- *Detection range:* This is the range at which tags or readers can be detected. Note that it does not necessarily mean that readers and tags can send or receive information. Thus, this range is the highest.

Although a variety of RFID classifications can be found in the literature, in [8] the authors classify RFID tags based on the computational power of the tags, because it is the most interesting from the privacy and security point of view. They introduce the following classification:

1. *Elemental or basic tags*, which are not capable of performing cryptographic operations such as generating random values or computing hashes.
2. *Symmetric key tags*, which are capable of dealing with symmetric key cryptography protocols. They are more expensive than the basic ones.
3. *Public key tags*, which are capable of managing public key cryptography protocols.

As RFID tags can be read by any reader in their cover ranges, some security and privacy issues must be taken into account. In fact, an eavesdropper could collect lots of information from citizens e.g. the brand and model of their clothes, the number of credit cards they have in their wallets, their mobile phones, the use of prosthesis or medicines, etc. Moreover, by making use of several readers strategically deployed in an area, it could be possible to track the locations of people. So, if proper solutions are not taken into account, people could be profiled and tracked.

The problems related to privacy can cause concern because a huge deployment of RFID technology could pave the way for a Big Brother effect, and there are more RFID tags in our environment than we would imagine. Hence, if no solutions are proposed, an attacker could attempt against our privacy. To give an example of an opposition to RFID deployment, in 2003, Benetton was boycotted when they tried to introduce RFID tags in their clothes [2]. After this, in the same year, several private organisations signed an agreement on how to use RFID technology in their products [5].

2.2.2 Security, Privacy and Scalability in RFID systems

The main advantage of RFID systems is that tags can be read without the need for visual contact. However, this advantage might also be a problem due to the fact that unauthorised people with the right equipment might be able to interrogate tags and obtain their information without being detected. This kind of unauthorised access might lead to the disclosure of confidential information.

With the aim to solve this problem, a wide variety of methods and protocols have been proposed. Designing private and secure, yet scalable, RFID identification protocols has been a big issue in recent years. The constrained computational resources on the tag's side makes this task very challenging. We have classified the current solutions as follows:

Killing: Electronic product code (EPC) tags address consumer privacy with a simple operation, which consists of tag killing. When an EPC tag receives a "kill" command from a reader, it renders itself permanently inoperative. To prevent wanton deactivation of tags, this kill command is PIN protected. To kill a tag, a reader must also transmit a tag-specific PIN (32 bits long in the EPC Class-1 Gen-2 standard). As "dead tags tell no tales", killing is a highly effective privacy measure. It is envisioned that once RFID tags become prevalent on retail items, point-of-sale devices will kill the RFID tags on purchased items to protect consumer privacy. For example, after you rolling supermarket cart through an automated checkout kiosk and paying the resulting total, all of the associated RFID tags will be killed on the spot. However, killing or discarding tags enforces consumer privacy effectively, but it eliminates all of the post-purchase benefits of RFID for the consumer. The receipt-less item returns, smart appliances, aids for the elderly, and other beneficial systems described earlier in this chapter will not work with deactivated tags. And in some cases, such as libraries and rental shops, RFID tags cannot be killed because they must survive over the lifetime of the objects they track. For these reasons, it is imperative to look beyond killing for more balanced approaches to consumer privacy.

Pseudo-identifiers: Even if the identifier emitted by an RFID tag has no intrinsic meaning, it can still enable tracking. For this reason, merely encrypting a tag identifier does not solve the problem of privacy. An encrypted identifier is itself just a meta-identifier. It is static, and therefore, subject to tracking like any other serial number. To prevent RFID-tag tracking, it is necessary that tag identifiers are suppressed, or that they change over time. Effacement of unique identifier neither eliminates the threat of clandestine inventorying, nor eliminates the threat of tracking. Even if tags emit only product-type information, they may still be uniquely identifiable in constellations, i.e., fixed groups. Use of random identifiers in place of product codes addresses the problem of inventorying, but does not address the problem of tracking.

Public-key encryption: Some of the solutions proposed try to encrypt the identifiers by means of public-key cryptography but, in fact, protocols based on public key cryptography, which are widely accepted in electronic commerce, banking, or access control, are an unrealistic option for low-cost RFID tags due to their limited computational power.

Distance Measurement: In [39], the authors demonstrate that the signal-to-noise ratio of the reader signal in an RFID system provides a rough metric of the distance between a reader and a tag. They postulate that with some additional, low-cost circuitry, a tag might achieve rough measurement of the distance of an interrogating reader. They propose that this distance serves as a metric for trust. A tag might, for example, release general information when scanned at a distance, but release more specific information, like its unique identifier, only at close range.

[63] proposes a privacy-protecting scheme that is called blocking. The scheme depends on the incorporation of a modifiable bit called a privacy bit into tags. A “0” privacy bit marks a tag as subject to unrestricted public scanning; a “1” bit marks a tag as private. It refers to the space of identifiers with leading “1” bits as a privacy zone. A blocker tag is a special RFID tag that prevents unwanted scanning of tags mapped into the privacy zone.

However, the blocker concept has limitations. Given the unreliable transmission of RFID tags, even well-positioned blocker tags might fail. Readers might evolve and exploit characteristics like signal strength to filter blocker signals [98].

Trusted Computing: The authors of [82] described an alternative approach to enforcement of privacy policies, such as those that rely on “privacy bits”. They describe how readers equipped with trusted platform modules (TPMs) can enforce tag privacy policies internally. Such readers can generate externally verifiable attestations as to their configuration in accordance with these policies. This approach does not address the problem of rogue readers, but it can facilitate or complement other forms of privacy protection.

Therefore, almost all efforts have been focused on RFID identification protocols based on symmetric key cryptography. Amongst all symmetric key identification protocols, the Improved Randomize Hash-locks (IRHL) protocol [64] is the most accepted due to its strong privacy and security properties, and its low computational requirements in the tag’s side (*i.e.* it only needs a pseudo-random number generator and a one-way hash function). In the IRHL protocol, for every tag’s interrogation, the reader generates a random number (*nonce*) r_1 and sends it to the tag. Upon reception, the tag generates another random number r_2 and computes the answer $a = h(r_1||r_2||ID)$, where ID is the secret identifier of the tag, $(\cdot||\cdot)$ is the concatenation operator, and $h(\cdot)$ is a one-way hash function. Finally, the reader receives the answer (a) and the nonce (r_2) from the tag. With this information, the reader (or the back-end) determines the ID of the tag by performing an exhaustive search in its database looking for an identifier ID_i such that $a = h(r_1||r_2||ID_i)$. When that happens the tag is identified as ID_i. Figure 2.10 shows a graphical description of this protocol.

Although the IRHL scheme is a private and secure RFID authentication protocol, it cannot be used when the system contains a large number of tags (*e.g.* like in manufacturing processes) because for every tag identification there is the need for an exhaustive search in the database. As a result of this lack of scalability, several

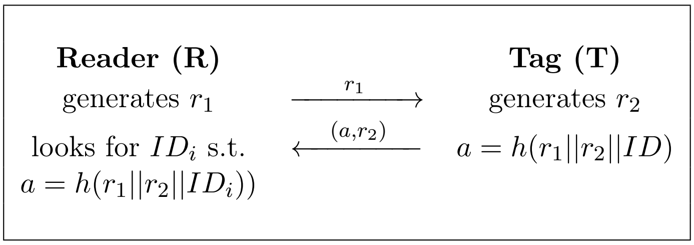

Figure 2.10: Improved Randomized Hash-locks Protocol [112].

protocols have been designed to reduce the linear complexity in the identification process of the IRHL protocol.

Tree-based protocols, such as the proposed by Molnar and Wagner (MW) [83], may be considered an alternative to IRHL in terms of scalability. The MW protocol achieves a time complexity in the identification process of $O(d \times \log_d^N)$, where N is the number of tags in the system and d is the branching factor of the tree that is used to store the tag's identifiers. When tree-based protocols are used, each tag stores a set of keys that uniquely identify it and, also, it must share, at least, one key with every other tag in the system. This sharing of private identification information might lead to undesired privacy leaks if a given number of tags are compromised[2]. The larger the number of compromised tags, the greater the risks for privacy [11].

Similarly to tree-based protocols, group-based protocols, such as the one described in [10], try to reduce the computational cost related to the secure identification of tags. In this case, tags are randomly assigned to groups. Consequently, each tag stores its own ID and the ID of the group to which it belongs. During the identification process, tags first send the group ID and then their own ID. In this way, the identification of tags is simplified because each tag has to be identified within its group instead of amongst all possible tags. However, group-based protocols could perform even better than tree-based protocols [10]. If a tag is compromised, the whole group to which the tag belongs is compromised too. Thus, from the scalability perspective, group-based protocols are very efficient, but with regard to privacy, they are not a good choice.

In the above protocols, the use of a single reader is assumed[3]. However, in 2007, Solanas et al. [108] introduced the idea of using multiple collaborative readers to make the identification process scalable whilst maintaining the high level of privacy of the IRHL scheme. Their proposal is aimed at efficiently identifying tags in applications where each tag must be continuously monitored whilst it remains in the system. This implies that readers must cover the whole system. Under this assumption, tags are constrained to move along neighbour readers[4] and therefore, neighbour readers therefore collaborate in order to guarantee efficiency during the identifica-

[2] Even a single compromised tag might lead to a privacy leak

[3] This reader be connected to a back-end that is responsible for the computation of the identification operations. Note that, readers are generally considered simple relays that forward identification information to back-ends.

[4] Two readers are said to be neighbours if their cover areas are not disjoint.

tion process. Efficiency is achieved by means of the so-called *reader's cache*, which is defined as a storage device in which a reader saves tags' identification data[5]. The protocol reduces the size of the readers' cache by considering that only the closest reader to some tag and its neighbours must store the identification information of this tag. By reducing the size of the cache, the identification procedure becomes more efficient. Despite the benefits in terms of computational cost provided by this protocol, assuming that readers are able to compute their accurate distance to tags is a bit unrealistic.

With the aim to solve the practical problems of [108], a novel protocol was proposed in [113], which aimed at identifying RFID tags efficiently without requiring special skills to readers (*e.g.* readers-tag distance computation). This protocol is based on the Solanas et al. protocol [108] but it has significant differences. A parameter $p \in [0, 1]$ is used to reduce the size of the readers' cache. The lower p the smaller the cache. Theoretically, when $p = 1$, the size of the readers' cache roughly equals the Solanas et al. protocol, whilst when $p = 0$, the minimum size is reached (at the cost of increasing the number of messages sent between neighbour readers). The idea is that p may be tuned up according to the application needs so as to favour the reduction of computations or bandwidth usage. Also, the protocol does not impose any constraint on the cover areas of the readers, neither on their communication architecture.

In the context of using multiple readers (connected to a centralised back-end), Fouladgar and Afifi [40] point out that, in many applications, tags are usually queried by the same set of readers. Therefore, they propose to cluster tags according to the readers that identify them more often. This idea improves the group-based proposals in the sense that tags are not randomly assigned to groups, but intelligently clustered according to the spacial location of the readers that identify them. By doing so, when a reader receives a tag's response, it first performs a search on the group of tags that it usually identifies. If it does not succeed, an exhaustive search is performed over the whole set of tags' identifiers. The problem of this proposal is that tags may have a long life-cycle and move through a wide variety of readers. In this scenario, the protocol could scale as bad as previous protocols based on symmetric key cryptography [64].

Recently, a new protocol based on collaborative readers was proposed by Trujillo and Solanas [114]. Similarly to previous protocols [108, 113], tags' data are stored on the cache of the readers that contain these tags in their cover area. However, this proposal is able to perform at least 50% better than the Solanas et. al. [108] and the Fouladgar and Afifi [40] proposals in terms of computational cost. The improvement is due to the use of a heuristic that allows readers to estimate the most probable previous location of tags. By doing so, once a reader could not identify a tag using its own cache, it asks for the information to another reader that is able to identify it with a given confidence. In addition to the reduction of the computational cost,

[5] This cache can be either an external database securely connected to the reader or a database internally managed by the reader itself.

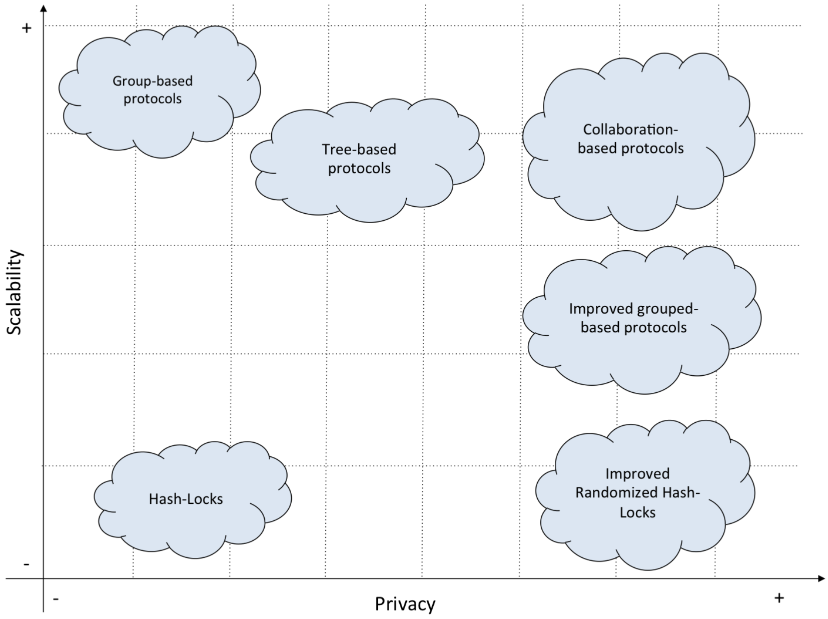
+
Group-based protocols
Tree-based protocols
Collaboration-based protocols
Scalability
Improved grouped-based protocols
Hash-Locks
Improved Randomized Hash-Locks
-
-
Privacy
+

Figure 2.11: Intuitive location of several identification protocols in the privacy-scalability plane. The top-right quadrant is where private and scalable protocols lay [114].

the proposal has significant improvements in terms of flexibility and usability with respect to other proposals like [113]. It is remarkable that readers do not need either to cover the whole system nor they have to rely on a neighbourhood relationship.

Some protocols are able to achieve good scalability and privacy by periodically "refreshing" the identification information stored in tags. These protocols add an "update" phase to the identification phase. Thus, the protocols assume that tags are able to change the information they store and recompute all necessary parameters of the protocol. Currently, this assumption is unrealistic for low-cost passive tags. Hence, we focus on the previously explained protocols and propose a new predictive protocol that improves the scalability of the aforementioned protocols whilst guaranteeing the same level of privacy of the IRHL scheme. Figure 2.11 depicts an intuitive distribution of the aforementioned protocols in the privacy-scalability plane[6].

2.3 Distributed Sensors

This section describes some fundamental concepts and ideas in which we found some of our proposals.

[6] This Figure is not intended for an exact evaluation but for an intuitive/approximative yet illustrative description of the relation between privacy and scalability of the aforementioned protocols.

2.3.1 Privacy as a product

Eric Houghes in [58] defines privacy in opposition to secrecy, as follows:

Definition 1. *Privacy is not secrecy. A private matter is something one doesn't want the whole world to know, but a secret matter is something one doesn't want anybody to know. Privacy is the power to selectively reveal oneself to the world.*

This intuitive definition and the very concept of privacy have received several interpretations in the course of human history. In the beginning, when the social interaction was basically analogical (i.e. prior to the digital age), as well as the assets and the available mechanisms, humans were trying to protect their privacy, either by trying to cover some asset, or hiding it. Privacy has been defined as a right quite recently. Its significant role as a human right became so important that due to the use of technological advancements from totalitarian regimes, it was included in the Universal Declaration of Human Rights.

The huge growth of the Internet, along with the vast amount of information that is becoming available to everyone, have created a new market. Systems that offer privacy to their users are becoming usual, and privacy is seen as a service for which users might pay.

The next steps in the evolution of the concept of privacy are the result of the current, wide use of Social Networks and Cloud services. Users can trade their privacy in exchange for free services. As a result, privacy is seen as a product or a good with which people can trade.

In this context, in [88], the concept of **Privacy as Product** (PaaP) was introduced by Patsakis and Solanas. The main contribution of that work was that it provided a new framework and a novel viable business model, in which people regain the control of their information and so that they are the only ones that can "selectively reveal themselves to the world", without any other further disclosures. Users may choose to receive money or another service in exchange for their information. Therefore, the paradigm of PaaP acknowledges the current situation that major social networks, search engines and cloud service providers have adopted, but tries to re-arrange the environment, returning a part of the revenues to their source – the users.

2.3.2 The Raykova-Vo-Bellovin-Malkin scheme

Private information retrieval (PIR) schemes allow users to export information stored in databases to an untrusted server, and allow to issue queries to the database without revealing the actual queries to the server, which does not neither know the records that have been returned as a result of the query. Even though the first two articles [27, 68] on PIR were very promising, as they showed that we can have solutions to this problem without having to replicate the whole database, up to today, we do not have any practical implementation, as the time and computational cost are prohibitive.

On a different line, fully-homomorphic cryptography, an encryption mechanism that was proved to exist quite recently [44], which could provide encrypted queries, proves to be inefficient for practical applications as well.

In the quest for efficient solutions to such problems, an interesting scheme, the Raykova, Vo, Bellovin and Malkin (RVBM) scheme [97], provides a feasible solution with small overhead [87] to an interesting and closely related problem to PIR schemes.

The RVBM scheme allows users/clients to anonymously perform queries to a server, without revealing any information about the query or the results, neither to the owner of the database nor to the server that hosts it. More interestingly, the server does not have access to the contents of the database, as all the contents are encrypted by keys selected by the users/clients.

The scheme has four main entities: the *User*, who supplies the data, the *Querier*, who queries the database, an *Index Server* (IS) which hosts the encrypted database, and the *Query Router* (QR), which performs as a proxy routing the queries from the Queriers to the Index Server, thus, providing anonymity to the queriers. In order to enable the aforementioned feature, the entities follow the procedure below, illustrated in Figure 2.12:

1. The user sends to the IS his/her encrypted data and a set of Bloom filters [16] that are used to manage keywords that he/she has in a dictionary.

2. The user sends a key to the Querier.

3. The querier encrypts his/her query with the key (received from the user) and sends the result to the QR.

4. The QR re-encrypts the query and forwards it to the IS.

5. The IS performs the search on the Bloom Filters and sends the results (encrypted) to the QR.

6. The QR replies to the Querier with the re-encrypted results.

These steps allow the Querier to identify which records in the database contain a certain keyword, without revealing any information to the IS. Moreover, the scheme can be further extended to provide aggregation queries instead of just keyword searches.

2.3.3 m-Health

With recent advances in wireless sensor networks and embedded computing technologies, miniaturised pervasive health monitoring devices have become practically feasible. The m-Health field has emerged as a sub-segment of e-Health (i.e. the use of information and communication technology (ICT) for health services and information) and it could be considered as a new way to access to healthcare services with specific capabilities of delivering real time information on demand.

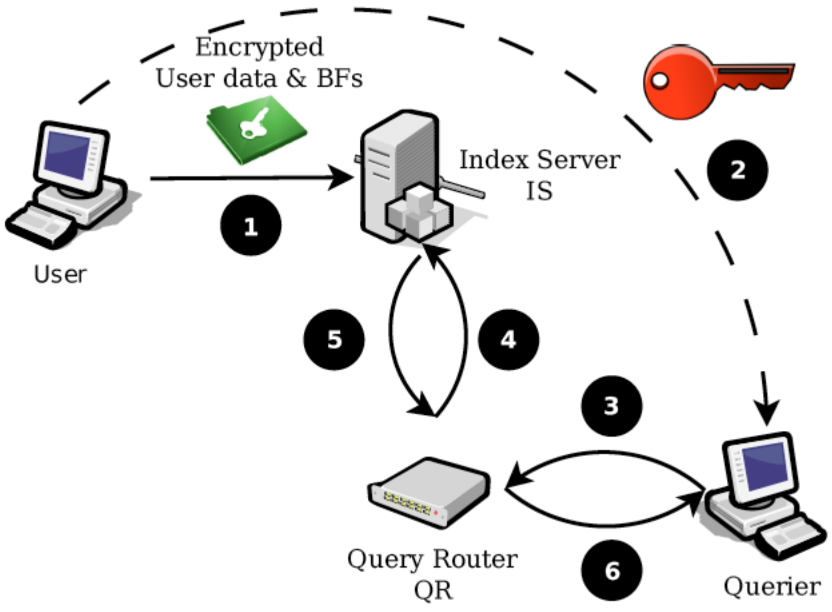

Figure 2.12: Raykova-Vo-Bellovin-Malkin scheme (taken from [88]).

m-Health applications include the use of body sensors connected with a mobile device that collects biometric or biomedical data; delivers the information to practitioners, researchers, and patients; monitors vital signs of patients in real-time; and provides direct care [45].

According to [106], m-Health redefines the healthcare services in three main aspects:

- m-Health allows the easy access to healthcare services and knowledge. Thanks to the inherent ubiquity of mobile devices, services may be accessed everywhere, anytime. In addition, data could be collected more easily regardless of the location of patients.
- m-Health is user-oriented. Users play a key role in an m-health service. Services should be where the users are.
- m-Health is personalised. Users receive customised services, that fit their personal needs.

The number of available healthcare applications for mobile devices has been growing during the last years. Many efforts are being devoted to improve the current m-Health technology in order to find new ways to offer a high quality of life to individuals as well as saving costs within the health systems, by providing proactive medical care.

The amount of data, along with the need for backups and synchronization among different devices, coupled with the rise of Cloud services, indicates that these devices will use cloud storage services to store this type of information in the near future. However, since there are many issues related to privacy in the Cloud, storing such sensitive and private information may expose user's privacy.

2.3.4 Body sensors

Body sensors are a mature technology, which has been evolving substantially for more than a century (cf. Figure 2.13). One of the major influences in the past years is the growing pressure from the medical sector to reduce costs whilst maintaining or even improving the quality of care. A potential solution to this problem is real time and/or remote patient monitoring by using mobile devices connected to body sensors of body sensor networks.

Renewed interest in body sensors has grown in order to support new applications in athletics, healthcare, emergency response, and consumer entertainment. The emergence and widespread deployment of low-power wireless radios and low-power circuitry have made sensors smaller, more energy efficient, and connectible to other devices and networks.

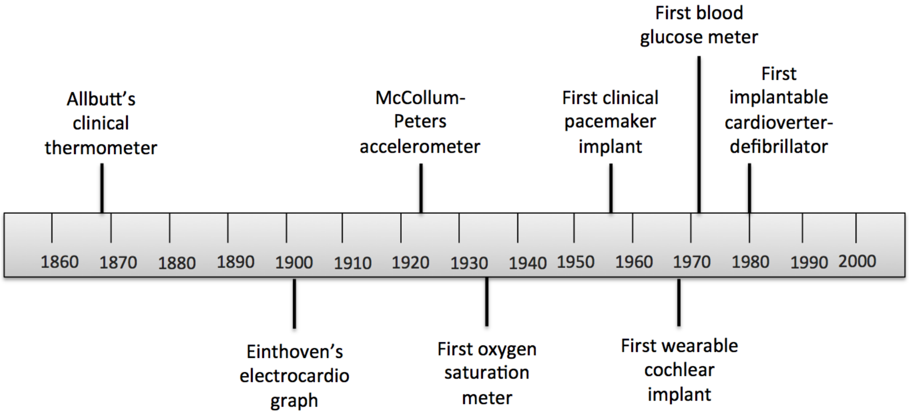
First blood glucose meter
Allbutt's clinical thermometer
McCollum-Peters accelerometer
First clinical pacemaker implant
First implantable cardioverter-defibrillator
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
Einthoven's electrocardio graph
First oxygen saturation meter
First wearable cochlear implant

Figure 2.13: Timetable with important innovations on body sensors during the last century.

Currently, smartphones can efficiently make blood tests with hand-held biosensors [7] with minimal cost. Sensors embedded in baby socks enable parents to monitor their baby's heartbeat and breathing rates [8]. Other devices[9] enable users to discretely track their heartbeat rate through bluetooth devices. Moreover, users are able to count their calories consumption[10].

[7] http://www.uri.edu/news/releases/?id=6369

[8] http://www.gizmodo.com.au/2012/11/these-heart-rate-monitoring-smart-socks-will-comfort_nervous-parents/

[9] http://www.polaraustralia.com.au/au-en/products/accessories/H7_heart_rate_sensor

[10] http://www.bodymedia.com/

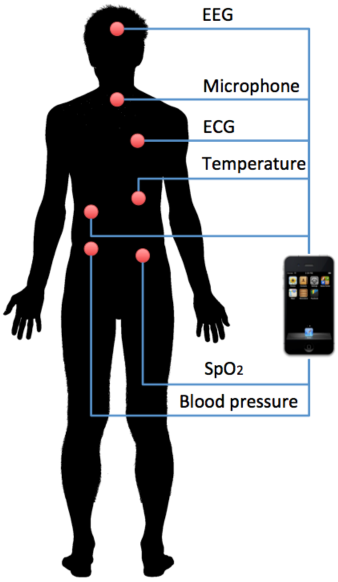

Figure 2.14: Brief overview of body sensors able to connect with a mobile device.

Since these sensors can be easily connected to smartphones (cf. Figure 2.14), it is very easy to get patients health data at the time when those data are gathered by the sensors. As a result, applications for healthcare are gaining interest among researchers (e.g. Research project: Optimized athlete body sensor networks for simulation-based performance analysis [11]).

In [106], the authors present a new concept called mCarer, a smartphone application and infrastructure able to monitor the movements of patients with different levels of cognitive impairments. mCarer can detect different states of alarm (previously defined by caregivers), based on the user's current location and the information about his/her health (e.g. if the user's heartbeat rate is anomalous, or the user is in a restricted area). This information is encrypted and stored in order to be analysed in case of emergency. Thanks to the roll access control used in mCarer, family members or carers are able to access the user's private information only under alarm conditions.

There are several examples of sensor integration within mobile devices. Trying to embed as many body sensors as possible in a mobile phone, LifeWatch [12] was developed and it includes sensors to monitor heartbeat rate, SpO2, glucose and even to perform an electrocardiogram (ECG).

Generally, body sensors are increasingly becoming network devices specially wireless, thus leading us to what we now call Wireless Body Sensor Networks (WBSN or just BSN). So multiple sensors are worn on the body, and each is capable of

[11] http://www.southampton.ac.uk/healthsciences/research/projects/optimized_athlete_body_sensor.page

[12] http://www.lifewatch.com/

sampling, processing and communicating wirelessly one or more physiological measurements or environmental parameters [86]. This can provide invaluable real-time data for analysis, diagnosis of health problems or even trigger security alerts.

The use of such networks may expose the security of their users in several ways [84]. However, we focus on the privacy. We can find more about BSN and their applications, mainly in healthcare, in [7, 56, 102, 118]. The typical uses are one-way, meaning that the user which is the source of the data, sends his data to the network and loses them. Of course he may select in many cases who has access to them, however they become an asset of the receivers, or to receive a service.

2.4 Smart Cities and Applications

2.4.1 Smart Cities

Due to the concept of "Smart City" is pretty new, as far as we know, there is no privacy protection systems to guarantee the privacy of citizens in the current literature, but some of the existing protocols applied in other environments can be adopted for an "Smart City", including some of the new protocols we present in this dissertation.

2.4.2 mHealth

Creative use of new mobile and wearable health information and sensing technologies (mHealth) has the potential to reduce the cost of health care and improve well-being in numerous ways. These applications are being developed in a variety of domains, but rigorous research is needed to examine the potential, as well as the challenges, of utilizing mobile technologies to improve health outcomes.

When patients suffering from a cognitive disease wander away and get lost, they may have serious or even fatal accidents. This usual behaviour, known as *wandering*, is a common problem related to people with dementia, from which over 40% get lost outside their homes. To avert this situation carers supervise patients and keep them safe. However, this often requires to lock doors or preventing the patient from leaving in other ways such as constant surveillance or putting them on drugs [75] [59]. Unfortunately, these methods limit the sense of freedom of patients and could have negative effects on their well-being. To mitigate these undesired effects, imaginative solutions, like installing a fake bus stop [20] have been proposed. However, these solutions are only applicable to very specific situations and patients.

With the aim to provide patients with more freedom and autonomy, Miskelly proposed in [78] the use of mobile phones equipped with GPS to continuously track patients, so that it was possible to locate and assist them if necessary. Before using the system, each patient gave written informed consent. However, the main problem encountered by the researchers was user compliance. The system proposed by Miskelly introduces the promising idea of using mobile technology but the lack of privacy related to the "Big Brother" effect [66] prevents this proposal from being fully accepted. Nevertheless, some commercial applications like the Columba bracelet [76] use this idea.

Instead of tracking patients constantly, Casas et al. [23] suggested the idea of using "alarms". A system of alarms only informs about the location of patients when something goes wrong. The alarm can be raised by patients (active alarm) or by wearable light devices when certain conditions are met (passive alarm). These conditions can be based on the location of patients (*i.e.* the alarm is triggered when a patient leaves a predefined perimeter, is near a dangerous area or is moving too fast) or based on the data of sensors, namely accelerometers [72], thermometers or heartbeat detectors [17]. Several commercial applications based on tracking and alarms are Urgentys [76], Simap [110] and GPS trackers [90]. These proposals, use both alarms and constant tracking. Note that although they use security and privacy methods such as deleting information, using Transport Layer Security (TLS), encrypting databases or using pseudonyms, none of them guarantee the privacy of the patients because users can ask for their location anytime, thus, deliberately invading their privacy.

In the current literature we can find solutions based on collaboration in order to face the "big brother effect" raised by these systems. Collaboration is paramount for many architectures that involve human beings. In [96], Ray et al. introduced the concept of awareness level as a measure of cooperation in the context of cooperative management and m-health, but there are other interesting examples of collaboration in the context of location-based services and tracking in [105], [108], [113] and [112].

Part I

Privacy in Location-Based Services

CHAPTER 3

Anonymity and Pseudonymity: A Distributed approach

In this chapter we propose a new location privacy preserving method — a distributed pseudonymizer. Our method is based on users' collaboration and solves most of the drawbacks of classic centralized pseudonymizers.

Contents

3.1 Introduction

Mobile devices having the capacity to locate themselves have gained importance in recent years. In the past, very few people had the chance to use those devices, but now, due to their relatively low cost, smart-phones and PDAs have become very popular. In addition, the launch of the iPhone, and all the clones that followed, have fostered the appearance of a large number of diverse location-based services (LBS) that allow the real-time follow-up of disabled or vulnerable people to guarantee their safety, help us find the closest restaurant or the cheapest gas station in our surroundings, and guide us along the best route to a given location. The great success of these services will promote the appearance of location-based providers that will be able to gather and analyse their users' location information. If the proper measures are not taken, the right of individuals to privacy could be endangered. As a result, the deployment of LBS could be significantly slowed down.

3.2 Our distributed approach

As we showed in the introduction chapter, TTPs are not a feasible solution, since they know all the information about the user. The privacy concerns in TTP-based schemes are exactly the same as in a direct connection between user and provider, because instead of trusting the provider we must trust the TTP, who can also be dishonest or it can lose the private data, due to an attack or just in an oversight.

Our aim is to design a new method that has the advantages of exact location methods such as pseudonymizers (i.e. simplicity and accuracy), and avoids their disadvantages (i.e. poor scalability and lack of privacy). Our idea is to replace the classic concept of pseudonymizer, understood as a TTP, by a distributed pseudonymizer consisting of a set of collaborative users. Our solution is based on the assumption that mobile devices are able to access the Internet[1] (i.e. they can use GPRS, UMTS, LTE, EVDO or WLAN), so that they can inexpensively exchange information. The communications between users are assumed to use the TCP protocol, so that we consider that messages properly reach each user. Also, we assume that we have a public key infrastructure (PKI), so that each LBS provider has a public and a private key.

In our model, we define three main actors:

- A set of users $\mathcal{U} = \{u_1, u_2, \ldots, u_n\}$ that cooperate to query a provider privately. Users are equipped with an IP enabled device.
- A server $\mathcal{S}_{\mathcal{L}}$ that is responsible for the maintenance of a list ($\mathcal{L}$) of IP addresses of collaborative users. Note that this is not a TTP because users do not share private information with it.
- A service provider $\mathcal{P}$ that answers queries from LBS users. $\mathcal{P}$ has a public key $PK_{\mathcal{P}}$ and a private (secret) key $SK_{\mathcal{P}}$, which are used to encrypt and decrypt sensitive information exchanged between the users and the provider.

Our scheme works as follows (cf. Fig. 3.1 for a graphical example):

Initially, a user u_1 that wants to obtain information from an LBS provider $\mathcal{P}$ has to obtain the list $\mathcal{L}$ from the server $\mathcal{S}_{\mathcal{L}}$[2]. Once u_1 has the updated list $\mathcal{L}$, she generates a packet with the following information:

Sour.	**Dest.**	**Query**	**Hops**
IP_{u_1}	IP_{u_2}	$E_{PK_P}(x_{u_1}, y_{u_1}, Q, k)$	R

Where IP_{u_1} and IP_{u_2} are the IP addresses of u_1 and u_2 respectively, (x_{u_1}, y_{u_1}) is the location of u_1, Q is the query, k is a random symmetric key and R is the number of hops (i.e. users) that will take part in the protocol. Note that the query

[1] Note that this is a realistic assumption because most if the time Internet connections are available from smart-phones.

[2] We assume there is some caching and updating technique for $\mathcal{S}_{\mathcal{L}}$ whose description is out of scope of this paper.

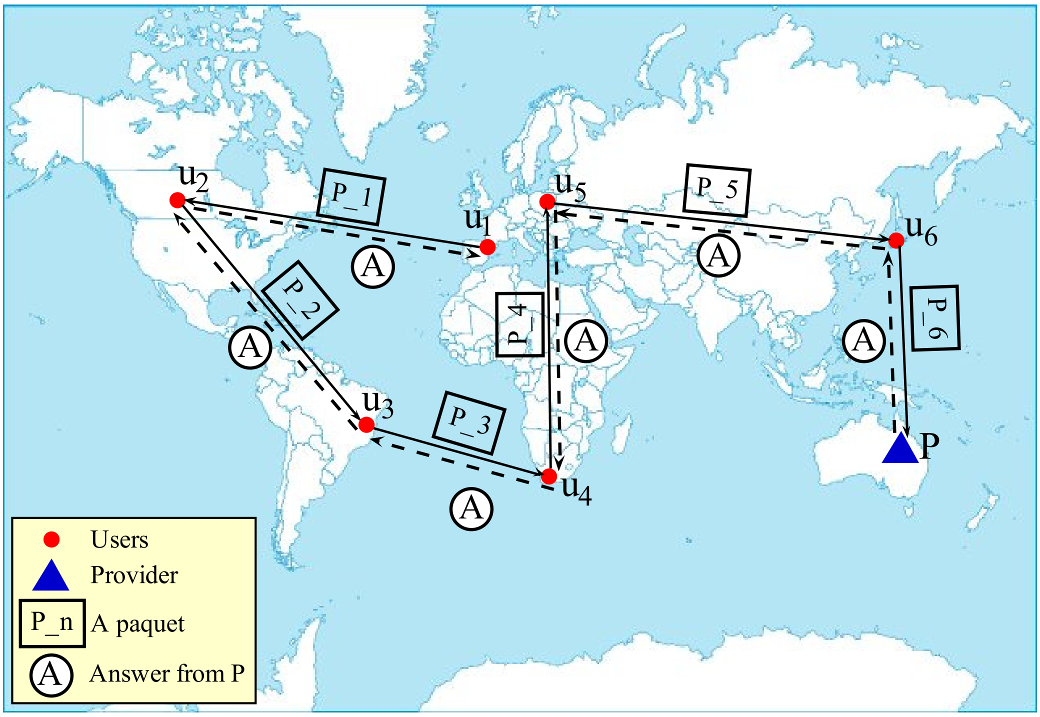

Figure 3.1: Communications example in the proposed scheme [93]

is encrypted with the public key of P. Moreover, u_2 is chosen randomly from $\mathcal{L}$ regardless of her location, which is unknown to u_1. Also, R is randomly chosen by u_1 so that u_2 cannot know whether u_1 is the source of the message or simply a collaborative user in the chain. After generating the packet, u_1 sends it to u_2.

When u_2 receives the packet, she retrieves $\mathcal{L}$, decrements the number of remaining hops, and generates a new packet as follows:

Sour.	Dest.	Query	Hops
IP_{u_2}	IP_{u_3}	$E_{PK_P}(x_{u_1}, y_{u_1}, Q, k)$	$R-1$

Like in the previous case, u_3 is randomly chosen from $\mathcal{L}$.

This procedure continues until the number of remaining hops is 0 (i.e. $R = 0$). In that case, the last user u_R sends $E_{PK_P}(x_{u_1}, y_{u_1}, Q, k)$ to $\mathcal{P}$. Then $\mathcal{P}$ decrypts the data by using his private key SK_P and builds a response packet containing the answer A to the query Q as follows

Source	Dest.	Encrypted answer
$IP_{\mathcal{P}}$	IP_{u_R}	$E_k(A\|\mathcal{H}(A))$

where $\mathcal{H}$ is a hash function, $||$ is the concatenation operator, and E_k is a symmetric encryption function with a shared key k.

The message from $\mathcal{P}$ is sent back to u_1 through the chain of users. Each user has to store the IP addresses of the users that are next to him in the chain of users, so that, she can forward the packets to the right users. The stored information can be organized in a resolution table:

Resolution table of u_1		Resolution table of u_2			Resolution table of u_R	
Prev.	**Next**	**Prev.**	**Next**	$\cdots$	**Prev.**	**Next**
$\emptyset$	IP_{u_2}	IP_{u_1}	IP_{u_3}		$IP_{u_{R-1}}$	$IP_{\mathcal{P}}$

When u_1 receives the answer, she decrypts it by using the key k and a decryption function D_k. The answer to the query is authenticated using $\mathcal{H}(A)$, taking into account that k was sent to $\mathcal{P}$ encrypted with PK_P. In addition, the location (x_{u_1}, y_{u_1}) that $\mathcal{P}$ sees does not belong to u_R, thus, $\mathcal{P}$ cannot profile u_R accurately nor effectively (i.e., the information he collects from u_R is useless), and at the same time u_1 receives a proper answer and her privacy is protected.

In order to make it clearer, let's consider a realistic case, based on Figure 3.1. Lets suppose a user called Alice who is in Barcelona (Catalonia, Spain) wants to submit a query to the LBS provider but she wants to protect her privacy by means of the distributed pseudonymiser proposed above. What she firstly does is to select a random user, called Bob, who is located in Canada. Note that Alice does not know the location neither the identity of Bob, and vice versa. Then Alice sends her query together with the location data, encrypted with the public key of the provider, to Bob. Due to the fact that the Bob's received message is encrypted, Bob is not able to read the content, so Alice does no require any kind of trust. Once Bob has received the message, he choose a new user, called Charles, in Brazil, and so on until the last user who is located in South Korea. When the last user receives the message he submits it to the provider, which is the only one able to decrypt the message and read the content, note that nobody in the chain knows the real source, even the provider is not able to infer if the message was submitted by a user located in South Korea, while Alice, the source of the message, is located in Spain. Once the provider answer the query, the response message is sent back to Alice through the chain of users.

3.3 Brief privacy discussion

Pseudonymizers are a very simple and clean solution to the privacy problem related to LBS. However, until now, they were based on intermediate entities that do not scale properly, are single-points of attack, are bottlenecks, and must be trusted by the users. Our solution distributes the task of the classic pseudonymizer amongst a set of users. Consequently, our approach scales better, is more difficult to be attacked because it is not centralized, it does not create bottlenecks, and no trust is required.

Privacy. Thanks to the use of public key cryptography, users can encrypt their locations and queries, so that only the provider can decrypt them. As a result, users do not need to trust each other — they just collaborate to propagate encrypted information. Even in the case in which several users collude with the service provider (i.e. by sharing the secret key SK_P of $\mathcal{P}$), they cannot be sure about who is the real owner of the location information because the number of hops R is randomly chosen and it is only known by the source of the query.

Authenticity and integrity. The provider $\mathcal{P}$ encrypts the answer using k, which can only be retrieved by those having SK_P. Hence, the strength depends on the length of k. A unidirectional hash function $\mathcal{H}$ has been used to ensure integrity.

Although we have overcome some of the shortcomings of pseudonymizers, our method and the former are vulnerable to the RSI and OI attacks [54].

3.4 Extended Protocol

As we have shown in the previous section, from a privacy point of view, the distributed pseudonymiser guarantees the privacy of the users, but some concerns about the network strength remain. An attacker is not able to infer the private data of the users, but he could cause a denial of service.

We detect two different DoS attacks, one of them consists on modifying the number of hops and the other appears when a user become non-collaborative.

3.4.1 Protection against number of hops modification

The number of hops that the query will make by the network before being sent to the LBS provider is selected randomly by the source user and only known to him. The random number will be chosen between a user defined margin, allowing the user to control the privacy level and speed of the answer:

- 0 hops, if the user does not want to protect his privacy, so that the response will be faster, or
- hops>0, if the user wants to conceal his identity and query, note that a larger number of hops generates a greater delay of the query.

Since the number of hops travelled is accessible in the packet, this could be a point of attack. A user who wants to deny the service could change the number of hops of the packet causing the packet to travel endlessly through the network without reaching its final destination. In this section we will propose a solution to this problem.

Our solution works as follow: Initially, a user u_1 that wants to obtain information from an LBS provider $\mathcal{P}$ has to obtain the list $\mathcal{L}$ from the server $\mathcal{S}_{\mathcal{L}}$ [3].

If the user already has $\mathcal{L}$, the latter is simply updated.

Once u_1 has the updated list $\mathcal{L}$, he generates a packet with the following information:

[3] We assume there is some caching and updating technique for $\mathcal{S}_{\mathcal{L}}$ whose description is out of scope of this work.

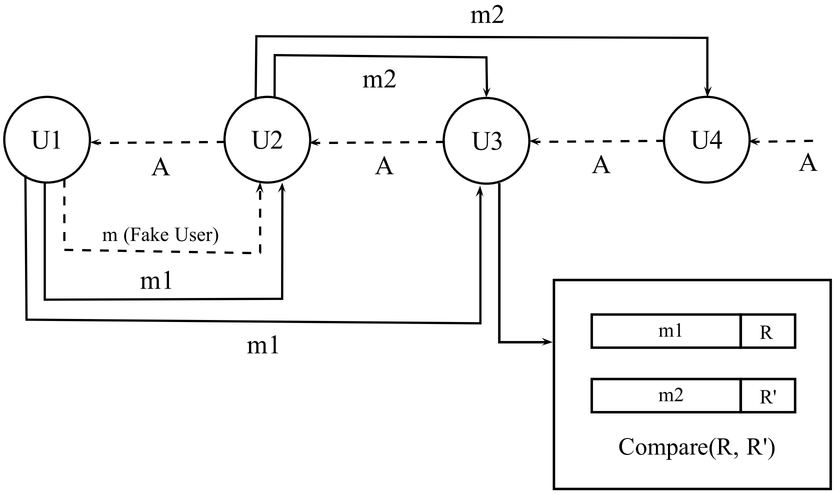

Figure 3.2: Users send the message to two next random users, doing so a user can check the validity of a received message.

Sour.	Dest. I	Dest. II	Query	Hops
IP_{u_1}	IP_{u_2}	IP_{u_3}	$E_{PK_P}(x_{u_1}, y_{u_1}, Q, k)$	R

where IP_{u_1}, IP_{u_2} and IP_{u_3} are the IP addresses of u_1, u_2 and u_3 respectively, (x_{u_1}, y_{u_1}) is the location of u_1, Q is the query, k is a random symmetric key and R is the number of hops (i.e. users) that will take part in the protocol. Note that the query is encrypted with the public key of P. Moreover, u_2 and u_3 are chosen randomly, from the users with the same reputation [4] as him, from the section of the list $\mathcal{L}$ sent by $\mathcal{S}_{\mathcal{L}}$ regardless of her location, which is unknown to u_1. Also, R is randomly chosen by u_1 so that u_2 and u_3 cannot know whether u_1 is the source of the message or simply a collaborative user in the chain. After generating the packet, u_1 checks that u_2 and u_3 are online and he sends the packet to u_2 and u_3.

When u_2 receives the packet, he retrieves $\mathcal{L}$, decrements the number of remaining hops, and generates a new packet as follows:

Sour.	Dest. I	Dest. II	Query	Hops
IP_{u_2}	IP_{u_3}	IP_{u_4}	$E_{PK_P}(x_{u_1}, y_{u_1}, Q, k)$	$R-1$

Like in the previous case, u_4 is randomly chosen from $\mathcal{L}$.
u_3 knows that he is the second destination, so he only needs to verify that previous user has not changed the package and takes appropriate action. If the previous user is not acting properly, he informs u_1 and $\mathcal{S}_{\mathcal{L}}$ that the user is dishonest. So the dishonest user will not be sent again by $\mathcal{S}_{\mathcal{L}}$ to users, and u_1 will send again the request to a new user.

[4] We understand reputation as a value obtained from the use of the system by a user, taking into account several parameters as contacts, time using the system, proper behaviour and os on.These parameters may vary according to the system.

This procedure continues until the number of remaining hops is 0 (i.e. $R = 0$). In that case, the last user u_R sends $E_{PK_P}(x_{u_1}, y_{u_1}, Q, k)$ to $\mathcal{P}$. Take into account that u_{R-1} doesn't need to send a copy of the packet to the next user after u_R, because there isn't a next user, he is the LBS. Then $\mathcal{P}$ decrypts the data by using his private key SK_P and builds a response packet containing the answer A to the query Q as follows

Source	Dest.	Encrypted answer
$IP_{\mathcal{P}}$	IP_{u_R}	$E_k(A \| \mathcal{H}(A))$

where $\mathcal{H}$ is a hash function, $||$ is the concatenation operator, and E_k is a symmetric encryption function with a shared key k.

As we can see in Figure 3.2, the source user of the query u_1 must create a fake user u_f before submitting the query, in order to send the query of the source user to u_2. Once submitted, u_1 will create his packet, and he will send it to u_2, so that u_2 can not deduce which is the origin.

Sour.	Dest. I	Dest. II	Query	Hops
$IP_{u_{Fake}}$	IP_{u_1}	IP_{u_2}	$E_{PK_P}(x_{u_1}, y_{u_1}, Q, k)$	$R + 1$

3.4.2 Unresponsive and non-collaborative users

With the previously proposed protocol we prevent a dishonest user from changing the number of hops R in order to cause a packet to travel through the network indefinitely, but the system is not protected from falls or non-collaborative users. What happens if a user of the chain loses network connectivity? In this case the components of the chain of messages would be left waiting for a response.

In order to prevent a user from waiting for a query indefinitely, we propose to use a time limit for each response, so once the time limit is finished, the user discards the expected response and he resubmits the message to a new random user.

Figure 3.3 shows how each user, after contacting the following user in the chain, waits a limit or Time To Live (TTL) until he receives the response message. If after that time the user does not get the response of the following user, he starts again the forwarding procedure. Doing so, we prevent a user from breaking the user's chain.

In order to keep the list of users free from non-collaborative users, we can add a reputation list, where non-collaborative users are banned. Each time a user requests a list of users to act as intermediaries, these users will be rewarded as participatory active users. If a user is detected as non-collaborative, it will be notified to $\mathcal{S}_{\mathcal{L}}$ and he will be penalised. Moreover, if a user repeats a misuse several times he can be banned from the service. On the other hand, it is possible that a honest user is not able to send back the response due some technical problem, then $\mathcal{S}_{\mathcal{L}}$ will be notified about the non-collaboration of this user, so he will be penalised but, statistically, if a user is honest he will have a positive reputation.

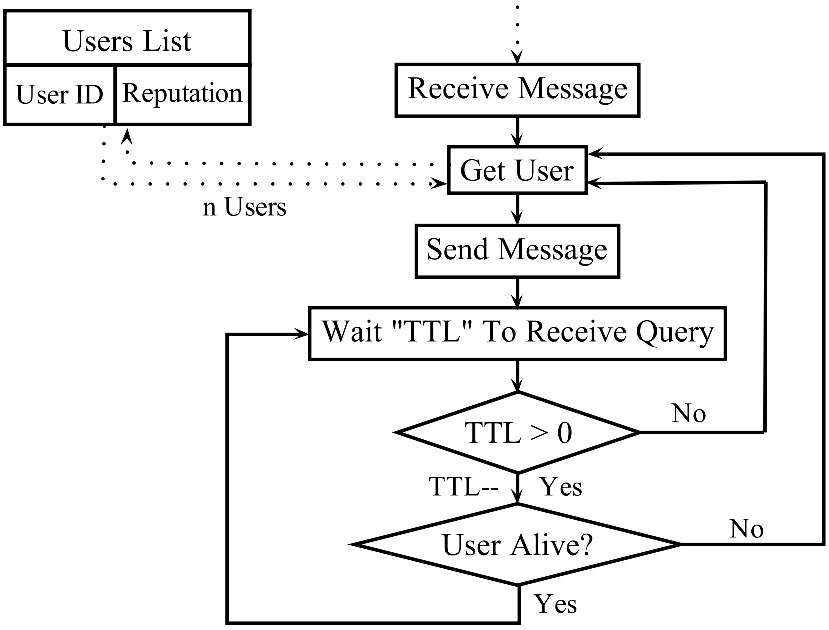

Figure 3.3: When a user receives a message, he asks for a new random user and submits again the message. If after a while the next user does not respond, he asks for a new random user and repets the steps until get a response.

3.5 Social Server List

We propose a server list which interacts with a *Social Network* for mutual benefit, where users have links with other users, and they can interact with the *Social Network*, as for example chatting or visiting profiles, to increase their reputation. In our proposal we do not need to create a new *Social Network*, since it can be applied over any existing *Social Network* (i.e. Facebook, Linked-In, Google+, etc.). Our protocol uses the structure of a graph for selecting the most appropriate candidate to send a message. To define the union among users in the graph, we need three main variables:

- **Common Friends:** Users can have relations between common friends, and this is the most frequent way to generate the union link in *Social Network*s, and it defines the weight of the links, in most cases positively, but in our scheme this is penalised because we want to send the messages as far as possible from our near contacts.

- **Interaction:** In most *Social Network*s users can interact with other users as for example visiting profiles, adding comments, chatting, etc., these actions will be used as a kind of union among them, in the same way as we do with the common friends, but in this case the interaction is considered a positive variable for the weight of the links. This is because using people with a high level of interaction we prevent the choice of unknown people that you added.

- **Collaboration:** Users can choose if they want to collaborate in the process of submitting a query to the location provider, so it is important to measure the collaboration of the users, to prevent that selfish users from benefiting from the server with no collaboration. The collaboration is the most important variable in our protocol, it is set as a positive value to increment the link between users in order to always choose collaborative users to submit queries. By doing so, non-collaborative users will tend to be left alone.

So, the weight of the links between users will be defined by the previous variables, in a way to discard non-collaborative users and prevent to send message to very near users [5], but encouraging users to use the *Social Network* to get stronger links.

In the next subsection we will discuss how to measure the different variables in order to generate the links between users.

3.5.1 Link Weight Meassurement

In order to combine the main variables to generate the links between users in the *Social Network* it is necessary to define the variables.

We normalise the values of **Common Friends** as follows:

$$\mathcal{CF} = \begin{cases} \frac{\mathcal{CF}^{max} - \mathcal{CF}}{\mathcal{CF}^{max} - \mathcal{CF}^{min}}, & if \quad \mathcal{CF}^{max} - \mathcal{CF}^{min} \neq 0 \\ \\ 1, & else \end{cases} \tag{3.1}$$

$$\mathcal{CF} = \begin{cases} \frac{\mathcal{CF} - \mathcal{CF}^{min}}{\mathcal{CF}^{max} - \mathcal{CF}^{min}}, & if \quad \mathcal{CF}^{max} - \mathcal{CF}^{min} \neq 0 \\ \\ 1, & else \end{cases} \tag{3.2}$$

The negative values are normalised following the equation 3.1 and the positives with the equation 3.2.

This value will decrease the weight of the links, so users will tend to use contacts with a low level of friends in common. By doing so, the queries have more probability of going far away from the net of friends of the former than choosing a close friend, who would send the message to another close friend, so finally the message would be sent by a close friend of the former.

The **Interaction** between users is measured with the numbers of events between users. Events can be defined by the *Social Network*, depending on the type of network, for instance, Facebook could use as events the fact of visiting profiles of other users or chat with them. We can normalise the **Interaction** value of users similarly to **Common Friends** values:

$$\mathcal{I} = \begin{cases} \frac{\mathcal{I}^{max} - \mathcal{I}}{\mathcal{I}^{max} - \mathcal{I}^{min}}, & if \quad \mathcal{I}^{max} - \mathcal{I}^{min} \neq 0 \\ \\ 1, & else \end{cases} \tag{3.3}$$

[5] Two users are called “near users” when they share a huge number of contacts.

$$\mathcal{I} = \begin{cases} \frac{\mathcal{I}-\mathcal{I}^{min}}{\mathcal{I}^{max}-\mathcal{I}^{min}}, & if \quad \mathcal{I}^{max} - \mathcal{I}^{min} \neq 0 \\ 1, & else \end{cases} \tag{3.4}$$

Negatives values of the interaction are normalised following equation 3.3 and the positives with equation 3.4.

When a user of the *Social Network* adds a new contact, the **Collaboration** variable is defined as "neutral", with a value of a 50%. This value will oscillate between 0% (minimun) and 100% (maximum) deppending on the degree of collaboration of users.

If a user U_0 sends a message to a user U_1, but U_0 does not receive any answer after a defined time $\mathcal{T}$, because U_1 is a non-collaborative user, then U_0 decrements the variable of **Collaboration** ($\mathcal{C}$) of user U_1, and sends again the message to a new user, otherwise if U_1 answers correctly, the variable is increased. In the case of U_1 was collaborative, but he did not answer due to U_2 did not answer him because he is the non-collaborative user, then U_1 also decrements the collaborative variable with U_2 and, statistically, after some iterations the weight of the link between U_0 and U_1 will become strong again, and U_2 will be marginalised for users.

Once we have all these data, we can define the **Link Weight** between users. We define this measure as:

$$Weight_{Link} = [(1 - \mathcal{CF}) + \mathcal{I}] * \mathcal{C} \tag{3.5}$$

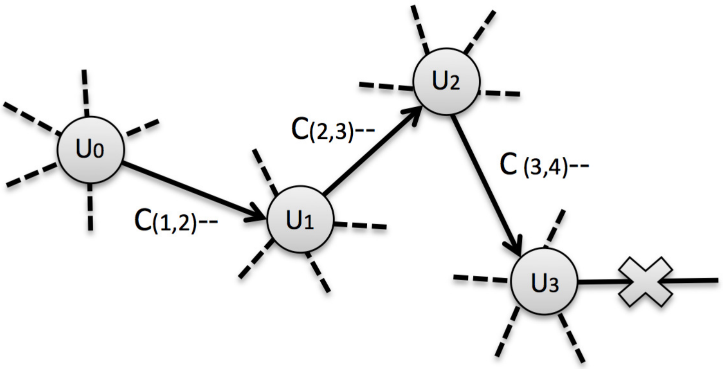

Figure 3.4: The collaboration variable (C_i, where i is the user at position i) decrements when users do not receive an answer from the next user. It can be due to another non-collaborative user in the chain of users, but all users are affected. Although, after some iterations, the last user (non-collaborative) will be alone, with no links, and other users will recover their links.

3.5.2 User Selection

A user can apply the Fitness Proportionate Selection over his graph of connections. This protocol is a genetic operator used in genetic algorithms for selecting potentially useful solutions for recombination. In Figure 3.5 we provide an example of the selected sample after following the next steps:

1. Sort the weights,
2. Compute the cumulative weights,
3. Pick a random number in $[0, 1] * TotalWeight$,
4. Find the interval in which this number falls into,
5. Select the elements with the corresponding interval,
6. Repeat k times.

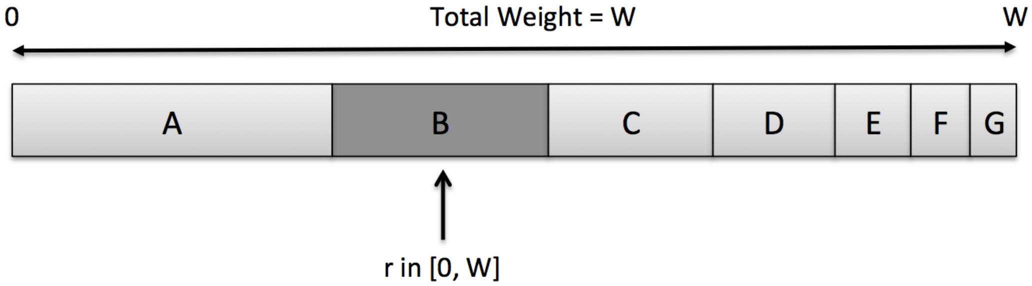

Figure 3.5: Fitness Proportionate Selection, where the sample B is chosen before applying the random function.

where N is the number of users linked to U and $f(i)$ the link's weight of the user i, the probability of being chosen is defined in equation 3.6.

$$p_i = \frac{f(i)}{\sum_{j=0}^{N-1} f_j} \tag{3.6}$$

This selection allows the users with a higher weight link to be selected with a high probability, but also it allows the worst users to be selected.

Similarly, when a user receives a query from a contact, he can execute the same function, but increasing the probability (eq. 3.6), to decide to collaborate or not. By doing so, non-collaborative users will not receive the collaboration of other users.

$$p_i = 2\frac{f(i)}{\sum_{j=0}^{N-1} f_j} \tag{3.7}$$

If a single user receives multiple queries at the same time, the queries will be sent forward taking into account the weight of links with the users who send him the query, from highest to lowest. So, if the user who send a query does not receive an answer, before restarting the process sending the message to a new user, he will wait a time interval $\mathcal{T}_i$ defined as follows:

$$\mathcal{T}_i = 2[t\Delta(h-i)] \tag{3.8}$$

where i is the user in the index i, and t is defined in equation 3.9.

$$t = Delay_{Network} + Delay_{Computation} \tag{3.9}$$

So, finally the total time of a query submission is:

$$\mathcal{T} = \sum_{i=0}^{h-1} 2[t\Delta(h-i)] \tag{3.10}$$

3.5.3 Attacker Model

We identify the following attackers and their dishonest intentions in our protocol:

- The service provider wants to obtain the maximum information possible about the users to perform a profile of them, but due to the fact that the query is submitted by a user who did not generate the message, the service provider is not able to link the query with the identity of the user who submitted it.

- A selfish user is a user of the system who wants to get a service but he is non-collaborative. This kind of user will tend to lose their links with other users, so they will not receive the service if they do not collaborate.

- A dishonest user tries to infer the queries of other users in order to track or profile them. In this case, dishonest users are not able to see the content of the queries, due to the fact that the content is encrypted with the public key of the provider, so only the provider can decrypt it.

- The server provider colluded with a dishonest user could be able to see the content of a message sent by a user, but the attackers can not know if the user who sent the message is the source or just a “bridge” user in the chain, since the number of hops is chosen randomly and only known by the source.

- A user who changes the content of the queries could perform a DoS attack on the service, but thanks to the use of cryptography, both the users and service provider are able to detect some modification in the content of the message, so users can decrease the weight of the link with the attacker. Also, this kind of attackers are difficult to find, since the contacts in the Social Network are known people by his contacts, and the reputation system proposed previously prevents the choice of an unknown or non-interactive contact.

- A man in the middle sniffing the data transmission of a single user, could be able to guess if a user is the source of a query, because he can see if a user sends a query when he did not receive any other query before. This kind of attacks may be prevented by using a secure channel (ie. https) through the *Social Network*, so the attacker could not differentiate the regular data transmitted by the *Social Network* and the user's query.

- The *Social Network*, it makes no sense for the *Social Network* to attack the system, because it already knows the private data of registered users, and it benefits from the proper functioning of the system.

Chapter 4

Query Privacy: Prime numbers and Homomorphic Cryptography

In this chapter we present a new protocol, which utilises some of the properties of public key cryptography and prime numbers, in order to achieve query privacy. This protocol can be easily adopted alongside some of the existing protocols as an extra privacy layer.

Contents

4.1 Introduction

Most of the privacy protection systems in LBS focus on hiding the users' identity, or obfuscating the users' location in some way, and only a few systems focus on hiding the query in order to prevent the users from being profiled. Therefore, we consider location services as a web search engine, that tries to learn about the queries of users.

Lets suppose a user wants to protect his privacy through basic public key cryptography. By encrypting his query with the public key of the provider, he can avoid an external party from inferring the query. However the user can also obfuscate the query in order to prevent the provider from profiling him. To do this, the user encrypts the real and fake queries in the same packet (instead of sending k encrypted messages):

$$\mathcal{E}_{PK_p}(q_0, q_1, ..., q_{k-1}) \tag{4.1}$$

where PK_P is the public key of the provider, q_i are the queries, and k is an integer positive number. If we assume a public key of 1024 bits, the user may achieve k-anonymity sending 1024 bits queries (including the $k-1$ fake queries plus the real one in a single message). Now let's assume n users are trying to submit one query each. They can achieve k-anonymity as in the previous example, that makes the protocol require a total of $1024 \cdot n$ bits. This cost may be assumable by the users, as they are just sending a single query, but the server receives $k \cdot n$ messages of 1024 bits each. This procedure results in data traffic n times higher than using a single query containing all queries from all users, plus n decryption operations, as queries are encrypted.

In this section we propose a new system able to guarantee the query privacy, even in the case of collusion among users and provider, with the same cost of sending single queries to the provider.

4.2 Basic Concepts of Cryptography

Before going further with the proposed protocol, it is necessary to know some of the primitives that we will use.

In the ElGamal [38] cryptosystem, in a group G, if the public key is (G, q, g, h), where $h = g^x$, and x is the secret key, then the encryption of a message m is $\mathcal{E}(m) = (g^r, m \cdot h^r)$, for some random $r \in \{0, \ldots, q-1\}$. The homomorphic property is then

$$\mathcal{E}(x_1) \cdot \mathcal{E}(x_2) = (g^{r_1}, x_1 \cdot h^{r_1})(g^{r_2}, x_2 \cdot h^{r_2}) = (g^{r_1+r_2}, (x_1 \cdot x_2)h^{r_1+r_2}) = \mathcal{E}(x_1 \cdot x_2)$$

Therefore, the product of two public key encryptions with ElGamal is equivalent to the encryption of the product of two messages.

4.3 System Model

Our system consists of:

- Service provider P, which has a public key PK_P and a secre key SK_P;
- A set of users $U = [U_0, U_1, ..., U_{k-1}]$ where k is a positive integer;
- A dictionary D consisting of a set of keywords *keywords*. Each keyword is linked with a point of interest (POI), previously defined, and known by the users. The keywords are a set of prime numbers;

In order to generate U, it is possible to make use of different systems, for example:

- Social networks: Most of the current social networks make use of the user's location in order to act as third parties and maintain communication among users. In our system, it is possible to make use of this feature in order to let the social network be the entity that provides us the set of users U in a specified space;

- Public WiFi access point: Many cities have developed free WiFi access points for their citizens. This connection can be used as a communication channel among the users of our service. The Wifi service provider would be able to generate a set of users connected to the same network, which would be a set of users sharing a location (i.e. the city or the Wifi antenna range);

- Ad-hoc networks (WiFi, bluetooth, etc): Most current smart-phones have internet capabilities, allowing for connections between users. By using this WiFi connection users may detect near users, inside the limits of their WiFi signals, and use them to generate U. Similarly, it is possible to make use of the bluetooth connection. In the newest version of bluetooth (bluetooth 3.0), connections can be established in a rang of 90 meters, equivalent to the range of WiFi, with speeds up 480 Mbps.

In our protocol, when k users want to send a query to P, queries are keywords linked with a prime number value. Each query is encrypted with PK_P and can be understood as $\mathcal{E}_{PK_p}(message)$, where $message$ is a prime number (i.e. 11: "hospital"). So, if we take the queries of the k users we have k encrypted queries, and by means of multiplicative homomorphism we convert the set of queries to a single encrypted query, where the message is the product of all queries encrypted with the public key of P. For example:

U_0: $\mathcal{E}_{PK_p}(3)$ (3: hospital)
U_1: $\mathcal{E}_{PK_p}(7)$ (7: pharmacy)
U_2: $\mathcal{E}_{PK_p}(11)$ (11: hotel)

P receives Q: $\mathcal{E}_{PK_p}(3 \cdot 7 \cdot 11) = \mathcal{E}_{PK_p}(231)$, and by means of SK_P it can decrypt the message and retrieve "231". Due to the fact that the number is the product of prime numbers, P can factorise it and retrieve the prime numbers which compose the message m : $231 = 3 \cdot 7 \cdot 11$ (hospital, pharmacy, hotel). Once factorised the message in its prime numbers, P can generate an answer for each query with just a single operation of encryption, $A = \mathcal{E}_{SK_p}(a_{hospital}, a_{pharmacy}, a_{hotel})$.

Our model offers a solution to the problem that appears in traditional methods; it reaches k-anonymity in the query with no need for false queries. So, we prevent the unnecessary computational cost that the server has to perform, and we allow users to make use of a paid service without extra charges from sending extra data.

Some current protocols try to achieve anonymity using fake queries [119], however, there are studies proving that in methods where the user generates fake queries, there exists a high probability to detect the false queries, by means of the study of user's patterns [89]. Since our method does not send false queries to the server, it avoids a point of attack, which most of the traditional methods do not prevent.

Our system is not limited to a single model, in this paper we propose two fully functional different models, with different advantages depending on the needs of user privacy (Table 4.1). The first model works as a distributed system, where the users collaborate among them in order to achieve a higher privacy. The second model is a centralised model, where a root user is responsible of the group's privacy, this model is most robust against errors.

More private	More secure
Distributed	Centralised

Table 4.1: Advantages of each model: The distributed model provides more privacy, but can easily suffer a DoS attack. The centralised model is more robust against these kinds of attacks, but the privacy level is lowest

In the next sections we will describe the two proposed models, and we discuss the cost and privacy for each of them.

4.3.1 Case 1: Centralised Model

In Figure 4.1 we see a centralised model, consisting of a group of users $U = [U_0, U_1, \ldots, U_{k-1}]$, where k is a positive integer and one of the users is the responsible for building the query by means of the product of all queries.

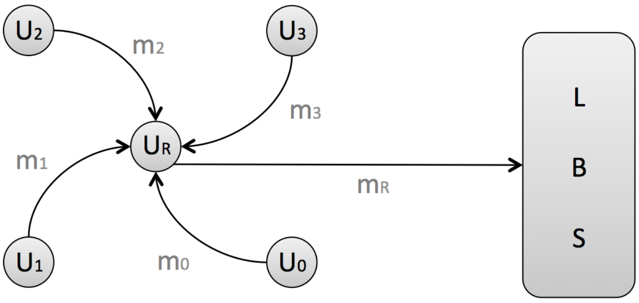

Figure 4.1: Query process in the centralised model, where all users forming the group send their encrypted queries to the root user, who is responsible for building the query consisting on the product of all queries.

m_0	$\mathcal{E}_{PK_P}(q_0)$
m_1	$\mathcal{E}_{PK_P}(q_1)$
m_2	$\mathcal{E}_{PK_P}(q_2)$
m_3	$\mathcal{E}_{PK_P}(q_3)$
m_R	$\mathcal{E}_{PK_P}(q_{0,1,2,3})$
$Answer$	$\mathcal{E}_{SK_P}(a_{0,1,2,3,R})$

Table 4.2: Generated messages, where m_i is the message generated by user i, and contains the query , $\mathcal{E}_{PK_P}$ and $\mathcal{E}_{SK_P}$ are the encryption/decryption functions keys of P, q_i is the query of user i, A is the message which contains the answer and a_i is the answer to the query i.

We assume that the service provider P has a pair of keys (PK_P y SK_P) and the system uses a known method to generate the obfuscation area, as for example in the proposed system in [109], in which a set of users collaborate among themselves to generate a centroid by means of privacy homomorphism and Gaussian noise in the locations, in order to prevent the users to see the location of other collaborative users.

In the centralised model, all users forming the k-anonymity send their queries encrypted with the public key of P to U_R. Once all the queries are received, U_R computes the homomorphic product with all of them, including his own query, so that U_R retrieves $\mathcal{E}_{PK_P}(q_{0,1,\ldots,k-1})$ and sends it to P.

$$\mathcal{E}_{PK_P}(q_i) \rightarrow P$$

Then, P, by means of its secret key SK_P, decrypts the message and it gets (q_0, q_1, ..., q_{k-1}) after factorising it. Due to each q_i representing a *keyword*, P is able to build and sign an answer $A = \mathcal{E}_{SK_P}(a_0, a_1, \ldots, a_{k-1})$ for each query, where a_i is the answer to the query i. Once the answer is built, P sends it to U_R, who broadcasts $(A - A_R)$ to all users of the group.

Thus, none user (except U_R) knows the existence of other users in the protocol, so they are not able to link the answers received with identities. They just know the existence of U_R, but due to the fact that A_R is not included in the answer, they cannot link U_R with an answer.

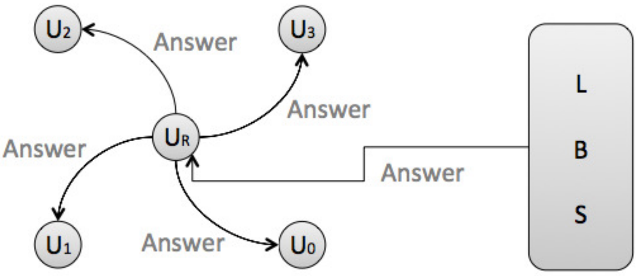

Figure 4.2: Answer process in the centralised model, where server builds an answer message with the answers of all the queries, sends it back to U_R, and he broadcasts the message to the whole group.

4.3.2 Case 2: Distributed Model

On the other hand, as we can see in Figure 4.3, our model is a distributed system consisting of a set of users $U = [U_0, U_1, \ldots, U_{k-1}]$, where k is an integer positive number and the generated messages for each user are reflected in Table 4.3. This method is proposed with the aim of avoiding the central user, of the previous system, to collude with the service provider.

As in the previous model, we assume that the centroid is done with existing protocols.

When a user U_0 wants to perform a query, he generates a first encrypted message with the public key PK_P of P. This message $m_0 = \mathcal{E}_{PK_P}(q_{f_0,0})$ contains the product of his real query q_0 together with a false query q_{f_0} chosen randomly from D.

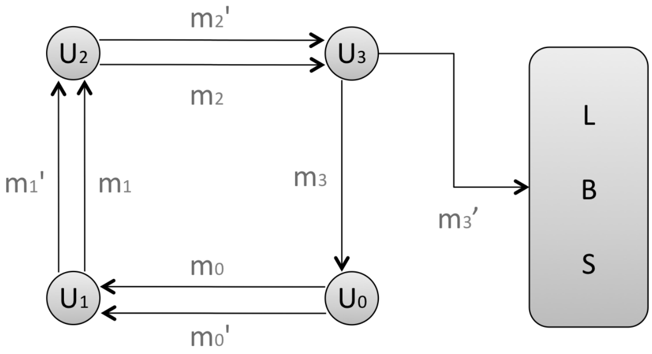

Figure 4.3: Distributed model with a double loop, where U_0 starts the query process with the message m_0, which is modified for each user until arriving at U_3 who sends the message m_3' to P.

Once m_0 is generated, it is sent to a nearby user U_1 together with the desired k-anonymity. U_1 cannot see the content of m_0, but he can perform a product over the message without the need for decryption by means of public key homomorphism. By using this property, U_1 is able to add new queries into the message, generating the message $m_1 = \mathcal{E}_{PK_P}(q_{f_0,0,f_1,1})$ that will be sent to a new user, and so on until it is received by the last user in the chain of users U_{k-1}, who will generate $m_{k-1} = \mathcal{E}_{PK_P}(q_{f_0,0,f_1,1,\ldots,f_{k-1},k-1})$.

m_0	$\mathcal{E}_{PK_P}(q_{f_0,0})$
m_1	$\mathcal{E}_{PK_P}(q_{f_0,0,f_1,1})$
m_2	$\mathcal{E}_{PK_P}(q_{f_0,0,f_1,1,f_2,2})$
m_3	$\mathcal{E}_{PK_P}(q_{f_0,0,f_1,1,f_2,2,f_3,3})$
m_0'	$\mathcal{E}_{PK_P}(q_{0,f_1,1,f_2,2,f_3,3})$
m_1'	$\mathcal{E}_{PK_P}(q_{0,1,f_2,2,f_3,3})$
m_2'	$\mathcal{E}_{PK_P}(q_{0,1,2,f_3,3})$
m_3'	$\mathcal{E}_{PK_P}(q_{0,1,2,3})$
A	$\mathcal{E}_{SK_P}(a_{0,1,2,3})$

Table 4.3: Generated messages, where m_i is the message generated by user i, and contains the real query with the false one, m_i' is the final query without the false query of i, $\mathcal{E}_{PK_P}$ is the encryption with the public key of P, q_i is the query of user i, A is the message which contains the answer and a_i is the answer to the query i.

In the case that a user wants a higher level of privacy than others and wants to prevent some type of collusion, it is possible to add $k' = max_{k-anonymity}/k$ false queries $q_{f_{[0,k-1]}}$ instead of just adding one, until the desired level of privacy is achieved. In this way, the total privacy of the user will be the lower between the both k-anonymities $min(k, k')$.

U_{k-1}, because he is the last user of the chain, must send m_{k-1} to the first user U_0, who started the protocol. In the same way you can apply a multiplicative operation over the content of the message, it is also possible to perform a division, so U_0 is able to remove q_{f_0} from m_{k-1}. Once the process is performed, the message is sent again by the user in the same way as before, and each user removes his false queries. Finally, when U_{k-1} removes $q_{f_{k-1}}$, the final message is $m'_{k-1} = \mathcal{E}_{PK_P}(q_{0,1,\ldots,k-1})$, that is the product of the real queries of each user, encrypted with the public key of P, so that only P can decrypt it by means of its secret key SK_P.

Once P has decrypted m'_{k-1}, it factorises the message, getting the queries like prime numbers (q_0, q_1, $\ldots$, q_{k-1}). Since each q_i is represented as a *keyword*, P could build the signed answer packet $A = \mathcal{E}_{SK_P}(a_0, a_1, \ldots, a_{k-1})$, where a_i is the answer of the query q_i.

Then, A is sent back to U_{k-1} (the user who sent the message m'_{k-1} to the server), U_{k-1} sends it to U_{k-2} and so on through the chain of users until the user U_0. In this way, users see all the answers of the queries, but they only know the identity of the previous user and the user next to them. Therefore, each user knows that one of the $k-1$ answers belongs to the previous or next users. So users achieve $(k-1)$-anonymity.

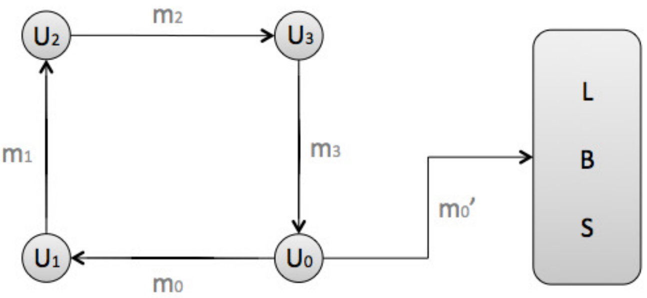

Figure 4.4: Query process in the distributed model with a single loop, where the user who creates the query is the responsible for sending it to the server.

With the centralised protocol users achieve the desired level of privacy even with the collusion of some users with the server, but in the case where users do not need a high level of privacy and they prefer a fast service, we can modify the protocol by having just a single loop in the chain of users. We can see a graphical example of the modified protocol in Figure 4.4, where with a single loop they can generate the same query, which means less time and bandwidth is required.

The modified version consists in a set of users, where the first user is the only one who adds false queries with the real one $m_0 = \mathcal{E}_{PK_P}(q_{f,0})$, the rest of the users just add the real query in the message until reaches the last user U_{k-1}, who returns $m_{k-1} = \mathcal{E}_{PK_P}(q_{f,0,\ldots,k-1})$ to U_0 in order to remove the false queries from the message and then sent it to P. Therefore, P receives a set of queries but it only knows the identity of U_0 and it will not be able to link the query with the user with a probability higher than $k-1$. Finally, P builds the answer message, sends back to U_0 and U_0 sends it to the other users, as we can see in Figure 4.5.

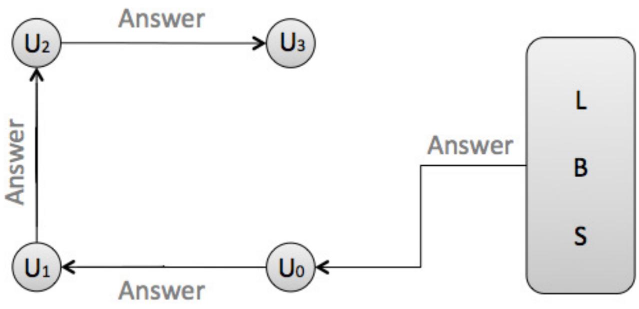

Figure 4.5: Answer process in the distributed model with a single loop

4.3.3 Study of k-anonymity and keywords dictionary

In order to perform the study of the k-anonymity, we assume a key size of 1024 bits (128 Bytes), big enough to guarantee privacy, but not too big to use it in a current smart-phone.

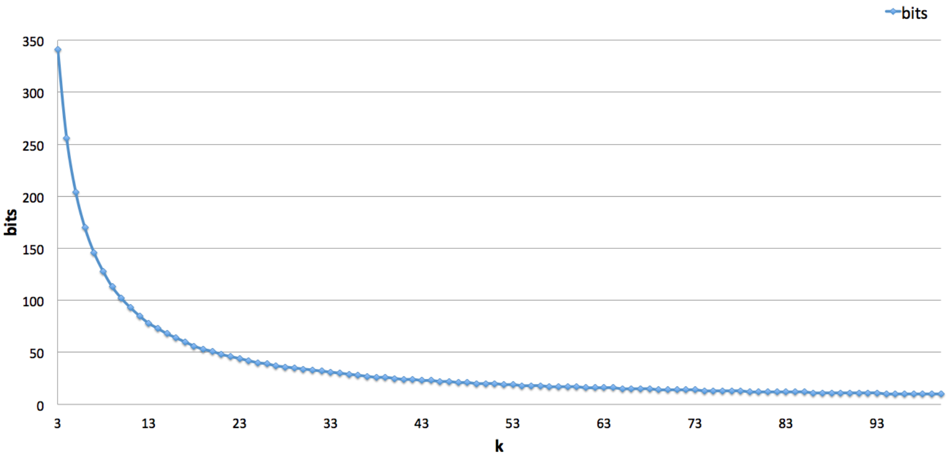

Figure 4.6: The chart shows the amount of *keywords* the system can handle according to the k-anonymity that we want to achieve. As higher is the k-anonymity, lower is the amount of words contained in the dictionary.

In order to enforce our protocol to achieve the users' privacy requirements, we should find the balance between the k-anonymity and the quantity of *keywords* forming the dictionary of the service. As we can see in Fig. 4.6, when the k decreases a little the available number of bits to represent the *keywords* grows exponentially.

Given a list of 1,000.000 primer numbers (used to represent the *keywords*), we obtain that the $keyword_i$ for $i = 1,000.000$ is $16,777.216$, so:

$$16,777.216^{42} \leq 1024bits \tag{4.2}$$

We can conclude that using a 1024 bits key, our service may be built with no problems, forming a dictionary of 1 Million *keywords* and a anonymity k of 42 users,

k-anonymity	*keywords*
100	172
75	1028
60	12251
50	82025
42	$>$ 1000000

Table 4.4: Values of the k-anonymity and keywords dictionary study with 1024 bits messages

which is likely enough for any service.

4.3.4 Cluster Heads

In the case that the group of users cannot satisfy the privacy requirements, due to there not being enough users to achieve k-anonymity, the system can make use of clusters in order to achieve the required anonymity.

Each group of users (cluster) will have a root user U_R (cluster head), who will be responsible for the communication with other clusters. Thus, the other users of the groups are completely unknown to the other groups, but they can exploit the advantages that other users offer.

By means of the communication among cluster heads and the use of privacy homomorphisms, it is possible to generate a centroid based on the location of the clusters. This centroid could be used to send the query to the server.

In this scheme, we compute the product of the queries in order to obtain a message containing the queries of all users in all clusters. Note that the maximum number of users ($k = 42$ taking into account our study) is a fixed value and it cannot be exceeded by the users of the clusters.

4.3.5 Factorising cost

For the simulations we have implemented a prototype of our proposal in Java, with a set of 100,000 *keywords* (more than enough for an LBS) and running in a Mac OS X (10.7.3), 1.7 GHz Intel Core i5 processor and 4 GB up 1333 MHz DDR3.

The first simulation consists of a set of users who receive an encrypted message with the public key of the provider and they compute the multiplicative homomorphism with the received message and their own query, previously encrypted with the same public key. The average time spent to perform the encryption of the query and compute the product is around 10 milliseconds. We then simulate the behaviour of the provider, who has to decrypt the received messages and factorise the content. These simulations have been performed with a set of messages containing from 2 to 50 queries (note that each query belongs a different user, so we simulated from 2 to 50 users). Figure 4.7 shows the results of the simulations, with a exhaustive search through the set of *keywords* to factorise. This factorisation can be improved

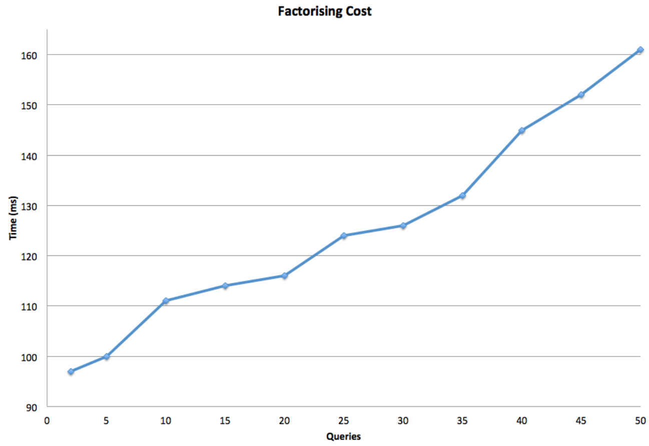

Figure 4.7: Time in milliseconds for factorising messages. The chart shows how the time to factorise grows almost linearly with respect to the number of queries to factorise.

in many ways, such as for example, building a multi-thread algorithm or sorting the *keywords* taking into account the probability of choosing each one.

4.3.6 Privacy Analysis

In this section we discuss how our system reacts against the types of attacks described above. For this we should analyse both swap protocols and proposed.

4.3.6.1 Attacker definition

As it occurs in all systems, it is not enough to guarantee the proper functioning of the system assuming only regular conditions, since it must also be robust against possible attacks, from both users of the system and external users.

In our scheme we can define 3 attacker models:

- **A user of the system U_s:** In this scenario the attacker is a user of the system, and he tries to modify or remove the messages of other users in the distributed model. These kinds of actions are known by DoS (Deny of service) attack.

- **The service provider P_a:** Usually the service provider is trusted, but it could be possible than a service provider can get private information from users. In this case, the service provider is able from seeing the content of the queries that users send to him.

- **An external user U_e:** Here, the attackers are not only the users of the system, but also external attackers, who try to infer the users' messages by means of an sniffer.

4.3.6.2 Privacy on centralised model

In the centralised model, except U_R, any user knows the identity of the other users that form the group. All the communication is done through U_R, so we prevent the attacks by users.

Instead, the system can be attacked by U_R, since he can try to deduce the content of a message by just encrypting all possible *keywords* with PK_P and comparing the messages. However thanks to the use of the ElGamal crypto-system, the same plain text is encrypted twice and will give different outputs. Alternatively, the dishonest user (U_R) can collude with P, since U_R knows the identity of each user, and can see the content of each query with SKp, which they can obtain from P. Therefore we recommend using some kind of system to select only trusted root users, like friends or contacts of a social network.

The advantages of the centralised model is the robustness of the system against DoS attacks. In this way, the centralised model works as an alternative for services that require more robustness than privacy, but with a minimum level of privacy.

4.3.6.3 Privacy on distributed model

In the distributed model, in the case in where U_i colludes with P, they could see $q_{f_0}, \ldots, q_{f_{k-1}}, q_0, q_1, \ldots, q_{i-1}$, where k is the number of false queries of U_0 and i the number of users. Here, each user of the chain would have an anonymity of (k+(i-1))-1 $= k + i - 2$.

If the colluding user was U_0, they could see the same as the server can see when it receives the queries, ie. the set of queries of all users $m = q_0, q_1, \ldots, q_{k-1}$. So, users get $(k-2)$-anonymity.

in the worst case, U_1 colludes with P to try to see the query of a user. By doing so, U_1 receives a query from U_0 and by means of the secret key of P that can see the query of U_0, but thanks to the use of false queries that U_0 adds to the message, U_1 is not able to distinguish which is the real one with a higher probability of $k+1$, where k is the number of false queries. So, U_1 and P can see $q_0, q_{f_0}, \ldots, q_{f_{k-1}}$, but they don't know which is the real one.

A single user can also attack the system without the need of collusion. This attack, as we previously defined, consists of encrypting all possible *keywords* of the system in order to compare the results with the message received from some user. We can avoid this kind of attacks with the use of the ElGamal crypto-system, since it is a probabilistic encryption.

The distributed model of our protocol is able to guarantee the privacy with an anonymity up to $k-2$ even in the case of collusion between a user and the service provider.

Chapter 5
W^3-Privacy: A holistic view

In this chapter we describe the three dimensions of user privacy in LBS. We introduce the concept of W^3-privacy, and we propose a new obfuscation method based on population density, anonymous payments and query uncertainty.

Contents

5.1 Introduction

As stated in 1, using LBS might be very beneficial but could lead to undesired privacy risks. Most of the current methods focus their efforts on location privacy, and they believe that protection that dimension is enough to guarantee user's privacy in location-based services. In fact, if the system does not protect all the dimensions of the users, the user information cannot be considered private.

In this chapter, we describe the three dimensions of user privacy in LBS and we coin the term W^3-*privacy*. Furthermore, we propose a method to obfuscate the location of LBS users. We show that our proposal achieves W^3-privacy when it is properly combined with anonymity and uncertainty methods.

Current methods consider that users are allowed to move freely, without restrictions, but in real live users cannot move freely over all space. They must use urban roads or streets, since people are restricted to walking or movement by vehicle. To achieve this balance, we have created a pattern of obfuscation based on high density locations over a city map. The LBS will know in advance the points of interest along the high level density locations, so users will be able to use one of these locations instead of their real locations.

5.2 W^3-Privacy

If we consider the privacy of LBS users we can distinguish three main independent dimensions, namely:

- *Identity privacy* (*i.e.* **Who** you are). Establishing a direct connection with the service provider, the latter knows the IP address of the former, which can be understood as a identifier.

- *Location privacy* (*i.e.* **Where** you are). This can be considered the main dimension and the most studied among the researchers working on the topic of privacy protection in location-based services, so that makes it the differential factor to be considered a location-based service.

- *Query privacy* (*i.e.* **What** you want to know). Even if a user protects his identity or location, the service provider, by means of OI or RSI attacks, can infer the identity of a user who asks regularly for the same or similar services. Doing so, it can profile and track the user.

We say that LBS is W^3-private if the service is given whilst the provider (i) does not know who the user is, (ii) where the user is, and (iii) what the user is asking for.

Most proposals concentrate on solving the privacy problems in a single dimension. In Section 2.1, we saw a survey of current proposals to protect the user's privacy, and as we can see most of these methods concentrate their efforts on protecting just one dimension whilst the other two dimensions remain visible for the service provider.

To address "identity privacy" most solutions rely on trusted third parties (TTP), that mediate between users and providers to hide their real identities using, for example, pseudonyms. Others try to hide users by means of onion routers [32]. In the first case, "identity privacy" is not guaranteed because the intermediaries know the real identities.

The most popular topic in LBS privacy is "location privacy". Several methods have been proposed to protect it, namely spatial and temporal cloaking (anonymisers and psudonymisers), individual location obfuscation, collaborative location obfuscation, etc. Most methods use TTP or collaboration to hide real locations, thus, they obtain location privacy at the cost of identity privacy. Only individual obfuscation methods guarantee location privacy without sacrificing identity privacy. "Query privacy" has captured the attention of the research community in the fields of Internet search engines and privacy preserving data mining. Several techniques to achieve query k-anonymity have been proposed and also private information retrieval (PIR) techniques are used. However, both approaches have serious practical limitations (mainly for computational reasons). Table 5.1 shows a brief summary of the privacy features of several methods to protect privacy in LBS.

Method	Who	Where	What
Pseudonymiser (TTP)		√	
Annonymiser (TTP)		√	
Individual Obfuscation	√	√	
Collaborative Obfuscation		√	
PIR		√	√
Our proposal	√	√	√

Table 5.1: Brief comparison of private LBS methods with regard to W^3-privacy [91]

5.3 Our proposal

In this section we describe our proposal to achieve W^3-privacy in LBS that require payment. First, we show how to guarantee “identity privacy”. Second, we describe how “query privacy‘” is achieved, and finally present a new obfuscation method to protect “location privacy”.

5.3.1 WHO

As we have previously introduced, a good way of protecting “identity privacy” is the use of onion routers [32]. Notwithstanding, this is not enough when LBS providers require the authentication of the subscribers. In these cases, an anonymous payment protocol has to be used.

We suggest using the partially blind signature of Abe-Okamoto [4]. Before sending queries to LBS provider, every user subscribes using the protocol described in [4], by doing so, users can authenticate without revealing their identity and providers can charge them for the service properly.

A partially blind signature scheme allows the signer to explicitly include common information in the blind signature under some agreement with the receiver. For instance, the signer can attach the date of issue to his blind signatures as an attribute. If the signer issues a huge number of signatures in a day, including the date of issue will not violate anonymity. Accordingly, the attributes of the signatures can be decided independently from those of the public key. We suggest the reader to refer to [4] and [117] for a better understanding.

5.3.2 WHAT

With the aim to preserve “query privacy”, users generate a number k of different queries $\{q_1, q_2, \ldots, q_k\}$. Thus, the provider is not able to determine which is the real query. This idea resembles the concept of k-anonymity and has been widely used to guarantee query privacy with Internet search engines [33].

5.3.3 WHERE

People living and moving in cities do not distribute uniformly. On the contrary, every city has streets, avenues and squares having a greater population density. Thus, it

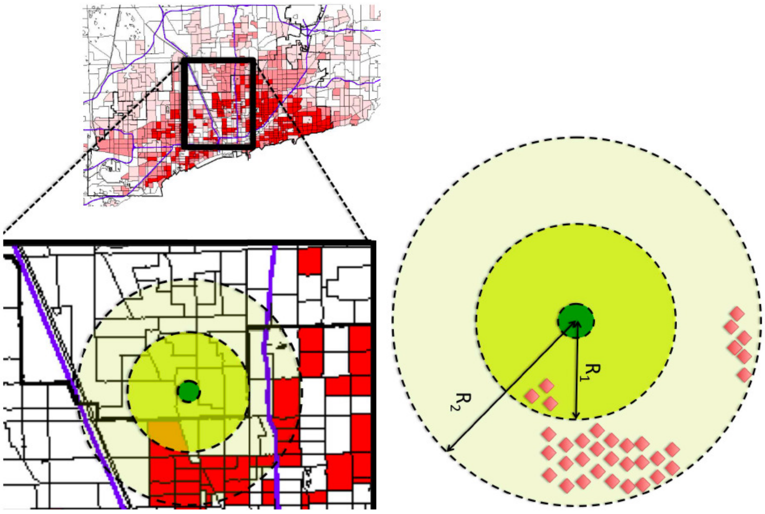

Figure 5.1: *(left)* Chicago's density map and user located on it. *(right)* Radius of uncertainty around the user and candidate points [91].

is feasible to obtain a density map such as the one shown in Figure 5.1(left). We assume that these maps are common to every LBS user and can be freely accessed. They can be understood as a set of points in $\Re^3$, (x, y, d), where x, y represent longitude and latitude, and d is the density in those coordinates. By using this information, users can determine the points/locations of maximum density, which are excellent candidates to be used as fake locations because they make difficult the object identification (OI) and restricted space identification (RSI) attacks.

To preserve their "location privacy" users define a series of uncertainty radius $\mathcal{R} = \{R_1, R_2, \ldots, R_m\}$ in which they try to find points with, at least, a minimum density d_{min}. These points could be used as fake points.

First, users locate themselves (*e.g.* with a GPS). Then, they select the area (defined by the radius) in which they want to find fake points. Finally, they use the information of the density map to determine the points that have a density greater than d_{min} that are within a range smaller than a given R_i. If no points fulfilling the required conditions are found, the user should relax the constraints and restart the procedure. Once the candidate points are found, users can select one of them randomly (alternatively, users can select the point which is closer to them, thus, reducing the error whilst maintaining the uncertainty).

5.3.4 Privacy protection on LBS

Privacy protection on location-based services is becoming a popular research topic, but we need to properly understand all the possible risks for the privacy of the users, as it may be violated, and that includes several factors to consider. The most popular factor is the identity of the user. Most privacy-protection schemes just face that dimension, and we may think that is enough to guarantee the privacy of the users, but actually it is not, as your interests or your location may be used to profile you, even when your identity is unknown.

Since privacy protection is understood as the protection of the user from being profiled by LBS, considering all the dimensions, a protocol that aims to guarantee the privacy of their users must face and guarantee the protection of the three dimensions. So, protocols that do not achieve this goal, should not be considered hundred percent secure.

Part II

Privacy in RFID and distributed sensors

CHAPTER 6

Smart Public Parkings

In this chapter we discuss how the ubiquitous use of information and communication technologies within the context of a Smart City might lead to transparent gathering of private data from citizens. We focus on transportation area and, more specifically, on the parking problems that might arise in big cities. We propose a set of procedures, based on privacy enhancing technologies, that allow the private, secure and efficient management of parking in Smart Cities.
The main goal of this chapter is to foster discussion about the privacy issues that might arise in a Smart City and to provide an example scenario (i.e. public parking) to demonstrate some interesting ideas and show some open problems.

Contents

6.1 Introduction

Countries are making great efforts to be competitive, attract investments and talent, reduce debt and be more sustainable. The struggle of countries for competitiveness has a smaller version in their cities, which are competing at an international level for investments, talent and quality of life, and they realise that the most promising path to success is the use of technology. Specifically, information and communication technologies (ICT) allow local governments and companies to develop ubiquitous innovative solutions that improve municipal operations in a variety of areas, such as transportation, energy, sustainability, e-governance, economy and communications.

In big cities, factors related to economies of scale help to reduce operational costs. However, managing big cities is challenging because the number of inhabitants grows steadily and the infrastructures and operational procedures have to be adapted to a growing and very demanding population.

In this context, local administrations have the need for smart procedures to improve the quality of life and the management of resources in cities. As a result of these needs, the concept of *Smart City* appeared and, although the idea of *Smart City* is pretty new, we can find several examples of cities that pursued this idea applied to a variety of areas (e.g. Amsterdam [70], Vienna, Toronto, Paris, New York, London, Tokyo, Copenhagen, Hong Kong or Barcelona [6]).

A very relevant area in every city is transportation. On the one hand, the management of public transportation is a very important and difficult issue that has been studied and companies, such as IBM, are proposing solutions to make it smarter [61]. On the other hand, private transportation is a cornerstone for the local government of any big city. The challenges related to private transportation are diverse, namely traffic jam management, tax collection, parking lots management, and so on.

In this chapter we enhance the definition of *Smart City*, which is a concept that has not been fully defined. We show the great advantages of *Smart Cities*, such as reduction of CO_2 emissions, improvement of the relations between citizens and administrations, increases in efficiency of public and private transportation, etc. However, we note that the easy gathering of data that occurs in ICT-based *Smart Cities* might open the door to privacy attacks from at least two sides: (i) from the infrastructure and (ii) from external attackers. To exemplify this situation we consider the special case of parking management within a *Smart City* and we describe a protocol that allows the private and secure management of the information required to control the payments in public parking lots.

The rest of the chapter is organised as follows: In Section 6.2 we provide some background on *Smart Cities*, we propose an extended definition for *Smart City* and we describe our case study. In Section 6.3 we describe our privacy-aware protocol for our case study and, finally, in Section 6.4 we briefly summarise its main properties from a privacy and security perspective..

6.2 Case Study

6.2.1 Smart Cities

In recent years many people have started to use the term "*Smart City*", but in many cases the meaning given to this term changes from person to person. Moreover, the term has gained a kind of marketing value that local governments want to benefit from. Thus, the definition of the term is frequently modified so as to adapt to the needs of the people using it in a particular situation. As a consequence, a number of different definitions and conceptual ideas regarding *Smart Cities* can be found in the literature.

From a very general perspective we could say that *Smart Cities* are those in which people can make their own choices and have a high quality of life combined with the efficient use of resources and the reduction of emissions. More specifically, a *smart city* considers six main areas/dimensions that are connected to the neoclassical theories of urban growth and development:

- *Smart economy:* Improve regional competitiveness and attract talent.
- *Smart mobility:* Improve the efficiency of public transportation and the management of private vehicles.
- *Smart environment:* Reduce the energy footprint and better use natural resources.
- *Smart people:* Promote human and social capital.
- *Smart living:* Increase the quality of life of citizens.
- *Smart governance:* Foster the participation of society and the interaction of the citizens with the administration.

According to Caragliu et al. [22] a city might be considered "smart" when it invests in human and social capital and traditional transport and in modern (ICT) communication infrastructures that fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance.

We complement this definition by adding that a city, to be "smart", should guaranty the privacy and security of its citizens so as to foster their participation and avert the *Big Brother* effect, which might raise concerns amongst privacy advocates.

To summarise, our definition of *Smart City* is as follows:

≪ ***Smart Cities*** *are cities strongly founded on information and communication technologies that invest in human and social capital to improve the quality of life of their citizens by fostering economic growth, participatory governance, wise management of resources, sustainability, and efficient mobility, whilst they guarantee the privacy and security of the citizens.* ≫

6.2.2 Public Parking

Parking a vehicle in a crowded city is a difficult task for several reasons. First, it is complicated to find a place and people waste a lot of time looking for a parking lot, and this time is translated to money due to the high cost of gasoline. Second, once a parking place is found, in many cases, drivers are required to pay some money depending on the length of time that the vehicle is parked. This payment poses several problems:

- **Need for change:** If the payment method is based on parking meters, drivers need to carry some change because, in general, credit cards are not supported. If the payment method is based on pay and display machines drivers might pay with money or credit cards, but they have to find the machine and then go back to the car to leave the ticket obtained.

- **Have extra-costs:** In some places prepaid RFID cards (wallet cards) can be used to pay for the service but they are usually not re-usable and drivers pay the extra-cost of the card every time they buy a new one [85].

- **Pay again to move the vehicle to another area:** In most cases, after paying for parking in a place, if the driver moves to another place (located in a different payment area of the city), he/she must pay again and cannot use the ticket previously issued even if it has not expired.

- **Renew the ticket:** When pay and display machines are used, if the ticket expires drivers must go back to the machine, buy a new ticket and leave it in the car. This is a very inconvenient procedure, specially if the driver is far from the parking place.

In the context of a *Smart City* we assume that a number of RFID readers are deployed so as to identify vehicles and control their payment status. Thus, in addition to the previously stated problems we identify some attacks against the privacy of the users that can take place and should be avoided.

- **Attacks from the infrastructure:** Some of current parking systems use contactless technology, but in most of cases users use an ID. If vehicles are identified with a single ID (e.g. the licence plate, or the like), the infrastructure can obtain a record of the locations that a given driver visits and can obtain extra-information that might endanger the privacy of drivers, namely their habits, their place of residence, their place of work, etc...

- **Attacks from external attackers:** If RFID technology is used inappropriately, external attackers could obtain the identification of the vehicle and clone it so as to avoid payment by stealing the identity of legitimate users.

In our case study we consider all these problems related to the payment and identification of the vehicles. We do not consider the problem of finding a parking place because it has been widely studied and several solutions already exist [69]. Thus, to address the aforementioned problems we need to design a procedure (or a set of protocols) that guaranties the following properties.

- **Anonymity:** Payments should be anonymous so as to avoid the identification of the user by the infrastructure and avoid undesired profiling. As we showed in Part I, by sending their identifications, users can be profiled by the service provider, as the service provider is collecting all the user's information, and thanks to the identifiers, it can link the information with an idendity.

- **Remote payment:** Payments might be done remotely without the need for change and without the need for returning to the vehicle or the parking meters. In a world where we can pay almost everything remotely, we should develop a remote payment method for parking meter to be considered "smart".

- **Transparent multi-area parking:** If users have paid for a given parking time and they change the location of their vehicle, they should be allowed to use the remaining time that they have (if any) in the new parking place. It is logical to believe that if you pay for parking in a city, you would be able to park everywhere inside the city, however most current parking meters make users pay again if they move their cars.

- **Untraceability:** External attackers and the infrastructure should not be able to distinguish two different payments from the same user. Thus, they cannot infer the habits or the places frequently visited by users.

6.3 Protocol

In this section we describe our protocol, which uses off-the-shelf privacy-enhancing technologies to address the problems identified in the previous section. We assume that users/drivers have an RFID card and a mobile phone that can communicate with this card. First, we describe the procedure to anonymously pay by means of e-cash. Then we describe how to use our protocol within the context of a *Smart City*.

6.3.1 Anonymous Payment

With the aim to break the link between the identity of the user and the payment he/she makes, we propose the use of anonymous e-cash. To obtain e-cash and proceed with the payment, users operate as follows:

1. **Get e-cash**: A user U_1 gets e-cash (electronic cash) from a bank. To do so, one can use a lot of existing protocols, for example the system proposed in the patent [103], in which a user asks the bank for a given amount of money in the form of electronic cash, in Figure 6.1 we show an example with a user asking for \$50 in e-cash. To do that, the user sends a request for some quantity of e-cash to the bank, and the bank sends back the e-cash with the requested value. In this procedure the bank signs the money so as to guarantee its validity. By using this procedure double spending is averted by the bank.

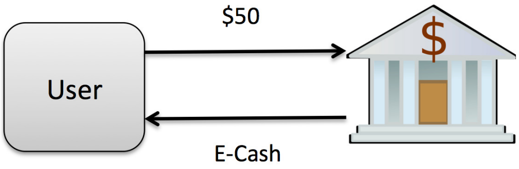

Figure 6.1: User asking to the bank for e-cash [92].

2. **Pay for the service**: When U_1 parks a vehicle in a public parking area that requires payment, he uses the previously obtained e-cash to pay for the

service by using a mobile phone (*cf.* Figure 6.3). To proceed, the user sends an activation message to the RFID tag located in the vehicle[1]. When the tag receives the activation message it generates a pseudonym using a one-way hash function h($ID_1||r$), where ID_1 is the private identifier of the tag, r is a random number generated by the tag, and ($||$) is the concatenation operator. Then, the tag sends the pseudonym back to the driver, who will use it to make the payment.

3. **Verify the payment**: Once the service provider receives the payment, it contacts the bank to check the validity of the e-cash. If the e-cash received is valid, the bank sends the money to the parking service provider (cf. Figure 6.2).

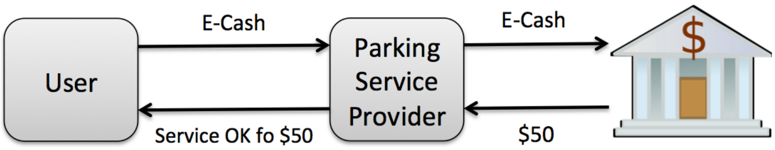

Figure 6.2: Graphic scheme of the payment and validation procedure [92].

4. **Determine and store expiration time:** The parking service provider converts the e-cash received into "parking time" and determines the expiration time for the user. Finally it saves this information in its database and informs U_1 about the expiration time.

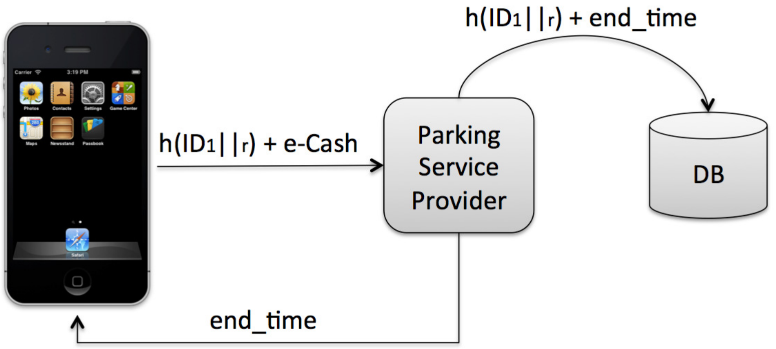

Figure 6.3: Payment by means of e-cash and a mobile phone [92].

6.3.2 Protocol Operation in a Smart City

We assume that the *smart city* in which our protocol is applied has a number of RFID readers deployed in public parking areas. Those RFID readers are able to identify RFID tags.

[1] This communication can be performed in a variety of ways, but the use of NFC is becoming popular and might be the standard in the near future.

- When a user U_1 parks in a monitored parking place, an RFID reader detects the tag in the vehicle and registers its current ID in its database DB. As stated before the ID of the tag that the reader will obtain is a pseudonym like h($ID_1||r$) that will be related to the payment issued by U_1. For the communication between the RFID reader (R) and the RFID tag of the user's vehicle (T), we can use the improved randomised hash-locks (IRHL) protocol [64]. As athours describe in [112], IRHL are computationally cheap on the tags side (they only need a pseudo-random number generator and a hash function). In the IRHL protocol, R generates a random number r_1 and sends it to T. Then, T generates another random number r_2 and computes the answer $a = h(r_1||r_2||ID)$ where ID is the secret identifier of T, ($||$) is the concatenation operator, and $h()$ is a one-way hash function. Finally, when R receives the answer (a) and the nonce (r_2) it determines the ID of the tag by performing an exhaustive search in its database looking for an identifier ID_i such that $a = h(r_1||r_2||ID_i)$. When that happens the tag is identified as ID_i. Figure 6.4 shows a graphical description of this protocol.

- U_1 proceeds to pay by means of his mobile phone, as we stated before (cf. Figure 6.3).

- Once U_1 has paid, the parking service provider updates the information about the expiration time in the database.

By using this procedure, users can pay anonymously, they can move from one area to another without paying again (because a centralised database contains all the information), they can extend their parking time by using their mobile phone, and they are not traceable thanks to the use of pseudonyms that can be changed every time a user parks a vehicle in a new location.

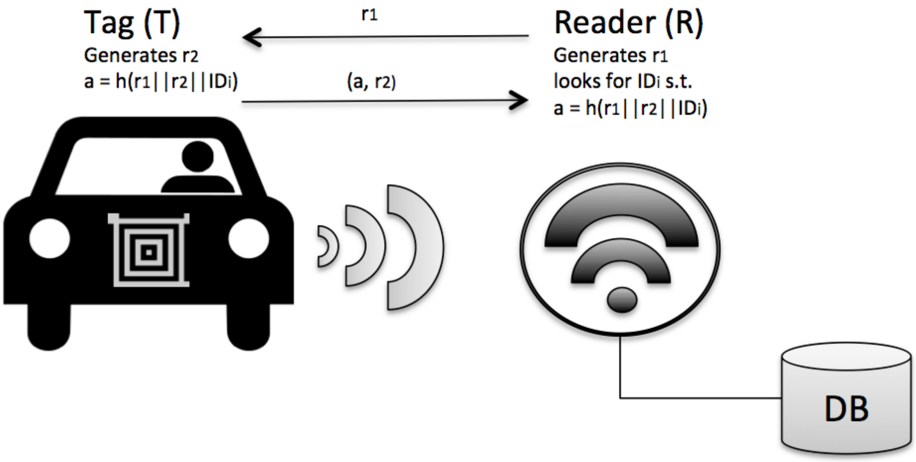

Figure 6.4: Secure communication between a vehicle and an RFID reader using IRHL [92].

6.3.2.1 Model 2: Cities with no adoption of sensor networks

In cities where there are no sensors distributed along it, normally there are some human reviewers who walk along the parking places verifying that all parked cars have paid for the services. In our system the reviewers can read the RFID tags of the parked cars through a RFID reader, so they can check if each car has a valid payment.

In this model, unlike the previous one, the system does not instantly detect the users when they are parking, so the delay for making the payment is not needed. Lets see the procedure:

- When a user U_1 parks in a place the system does not realise it until a human reviewer checks the car. So U_1 will pay before being identified by the service.

- Once U_1 has paid, the service adds the information to its database, so it adds ID_1 and the end time paid $end\_time = current\_time + paid\_time$.

6.4 Discussion

We have designed our protocol so as to provide anonymity, remote payment, transparent multi-area parking and untraceability. Next we briefly discuss why our protocol achieves these goals and how its help to protect citizens' privacy within the scope of a *Smart City*.

- Anonymity: Thanks to the use of e-cash, the infrastructure cannot relate the payments with the identity of the user. This would be only possible if the infrastructure colludes with the bank. However, this does not seem to be a realistic scenario.

- Remote payment: As the payment is no longer linked to a machine or a parking meter, users are able to pay remotely using their e-cash through a mobile phone.

- Transparent multi-area parking: Information about payments and identifiers are stored in a centralised database. Thus, if users change the location of their vehicles, they do not need new tickets.

- Untraceability: At any moment, the user might decide to send an activation message to the RFID tag of the vehicle. By doing so, the user makes the tag generate a new pseudonym that will be used to pay and to identify the vehicle. Due to the fact that pseudonyms change, it is impossible for the infrastructure to trace users by means of their ID.

From a security point of view, the use of the IRHL for the communication between vehicle tags and infrastructure readers guarantees that the pseudonyms generated by tags cannot be cloned. Thus, the security of the drivers is also guaranteed. The down side of IRHL is its computational cost on the readers side, however, it has been shown that it is possible to obtain efficient identifications by using the collaboration of multiple readers (which would be highly applicable in a *Smart City* scenario) [112], [114], [113].

Chapter 7
Cloud and Body Sensors

In this chapter we propose a privacy-aware protocol for body sensor network (BSN) in the cloud that guarantees the right of patients to privacy with an economically sound solution. The new twist that the proposed model provides is that a part of the revenues is returned to the source of the information, the users. We leverage the recently proposed concept of Privacy as a Product (PaaP) in combination with the Raikova-Vo-Bellovin-Malkin (RVBM) protocol to develop our scheme and we show that it guarantees privacy and it is viable from a technical and economical perspective.

Contents

7.1 Introduction

Body sensors have been around for a long time. Since the invention of the clinical thermometer in the 19th century, the evolution of body sensors has accelerated exponentially. The development of new materials, advances in miniaturisation, and the use of Information and Communication Technologies (ICT) have elevated body sensors to a new level in which they are naturally integrated with humans. Thanks to the use of body sensors we can monitor a variety of body signals, namely temperature, heartbeat rate, blood pressure, blood oxygen saturation, glucose, etc.

Although body sensors were initially designed as individual measurement devices (see for example the thermometer), they can no longer be understood as isolated gadgets. Instead, current body sensors aim at collecting data in a collaborative way, since the data collected by all of them could be studied globally to obtain a more accurate picture of the actual state of the people wearing them. In addition, body sensors can create ad-hoc networks that allow them to communicate and exchange information amongst themselves and with other entities that might be used to collect and analyse biomedical and biometric information.

As a result of the evolution of body sensors from individual entities to complex, interconnected devices, concepts such as electronic health (e-health) and mobile health (m-health) have gained momentum. Nowadays, it is possible to remotely monitor the body signals of patients and introduce changes in their treatments or urgently take action in case of emergency. Also, the biometric data collected by those sensors could be crossed with other data such as location data, medication intake information, or even medical health records. Consequently, all this information can hardly be stored and analysed in locally managed systems. Instead, the cloud paradigm in which storage and computational capacity are leveraged from a third party appears as a very serious trend.

Figure 7.1 depicts a simple version of this paradigm shift: Users collect the data of their body sensor networks in a static or mobile communications device. The collected data is periodically sent to a cloud service provider (CSP) who is responsible for storing them and (optionally) could apply some transformations on the data for efficiency reasons such as compression or indexing. The cloud acts as a pool of data that can be accessed by other entities like hospitals and research centres for primary use. Those parties would have access to the data so as to monitor the health status of patients, control the evolution of a given disease, or carry out research studies. In addition, the CSP could achieve some revenues by selling sanitised (or raw) data to third parties like pharmaceuticals, biosensors manufacturers or, sports equipment retailers, which make secondary use of the data (i.e. they use the data not to provide healthcare services to the users but to obtain information that could help them to improve their products and increase their sales. Thus, they are ready to pay the CSP in order to get access to this data).

The simple cloud-based model depicted in Figure 7.1 has many benefits for the final user, some of them are the following:

- Availability: Cloud services are already a reality and can be easily accessed. The storage of information in the cloud is pretty simple and the sharing of this information in the cloud has become commonplace (see for example the case of Dropbox).

- Reliability: Due to the fact the cloud infrastructures are largely redundant, it can be said that a properly managed cloud is reliable. Thus, users can safely store their information, since a number of backup units take care of the proper storage of their data.

- Low-cost: As a result of economies of scale, CSPs can offer services at reasonable costs. In fact, the public cloud paradigm, in which people share resources such as CPU, memory, etc, is becoming very common and the cost for end users is decreasing steadily.

It is evident that using the cloud has a lot of benefits. However, cloud users are handing over their data to the cloud, and in some situations, sharing the raw data with the cloud could be seen as an important privacy risk. This is specially true in the case of data coming from body sensor networks, which can be considered highly sensitive personal data.

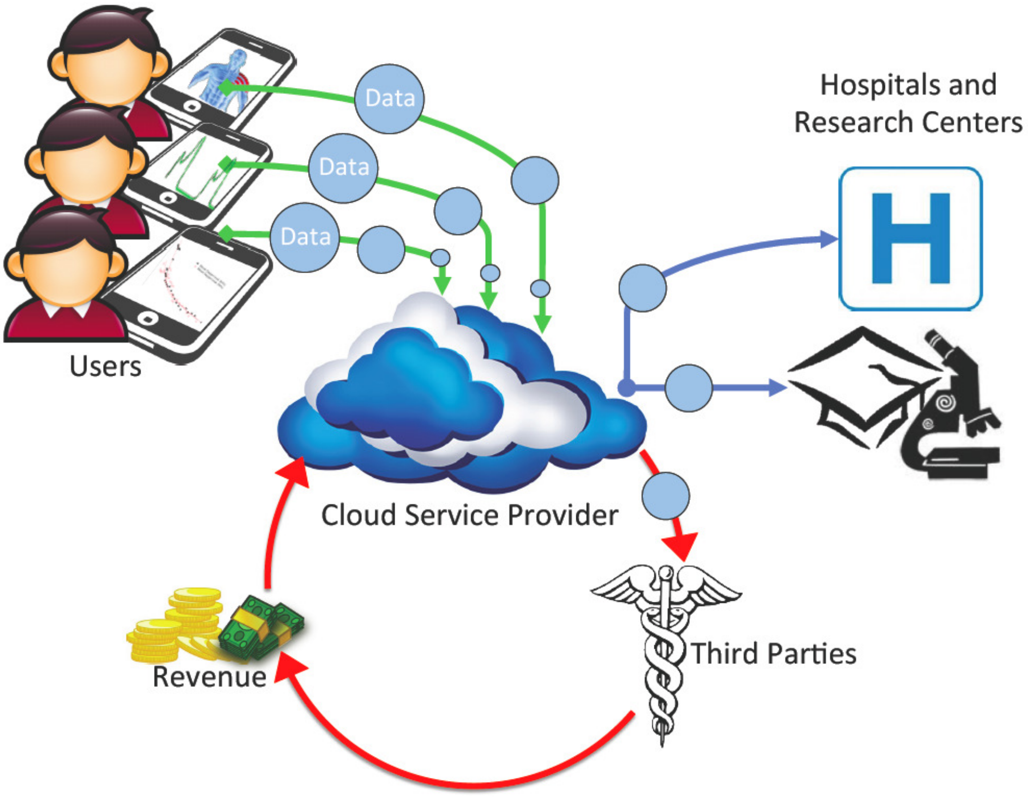

Figure 7.1: Simple model for cloud storage of data collected by body sensors. The data are shared with entities for primary and secondary use, and the cloud service provider could make revenue from this data sharing.

7.2 A private health center in the cloud

7.2.1 Main actors and desiderata

In this section we model our problem, state the requirements and provide the necessary structures and mechanisms that will build our framework to allow sharing of body sensor measurements with privacy in the cloud and to enable a novel business model based on the concept of PaaP.

The proposed model comprises four entities: the user Alice, the Cloud Service Provider (CSP), the Query Router and the Bidder. These entities could be mapped to the RVBM roles of user, IS, Query Router and Querier respectively.

Our first goal is to keep the stored information away from the eyes of the CSP, consequently, a user's data should be encrypted using a key that chosen by him/her. In order to make the provided solution efficient in terms of key management, we propose the use of a public key algorithm, which allows re-encryption, like El Gamal. This selection might slow down performance. However, it makes data sharing easier and more secure. Moreover, current smartphones and tablets can easily handle such processing cost.

Our second goal is that users could share their data without having to disclose any additional information to the server, nor to provide any additional key to the sharers. Since the data will be encrypted, querying the information will be very hard, therefore, we propose the use of Bloom filters, as in the RVBM protocol, but with two additional filters. Depending on the information that is stored in the files, we consider that the user might want to allow a reduced set of search operations regarding on the type of querier. The user might choose to allow querying different tags to the sharers and the queriers. Based on that, the user (Alice) creates three keyword dictionaries and the corresponding Bloom filters. The Bloom filters will be encrypted but with different keys in order to preserve their independence.

From the above, it becomes clear that Alice is in full control of her data, she will be the one to chose who has access to her data, for how long and the role of the CSP is just to host the information and allow the necessary tasks to be executed.

In our model, the Bidders have the chance to bid for subsets of the information that the user is sharing and willing to trade. However, in order to do that the Bidder must be able to certify that he is in fact a bidder. The certification of the bidders is a task that is made by the Query Router (QR). The QR not only forwards the queries anonymously, but acts as a certification authority for the Bidders, granting them the appropriate credentials to place bids.

7.2.2 Managing and trading the data

Whenever Alice wants to upload a file, she picks a key and sends the encrypted file to the CSP, along with a Bloom filter for her searches and one more if she wants to allow sharing. All the filters that Alice creates are sent encrypted to CSP and each of them with a different key, selected by her. To revoke the rights of a user, all that Alice has to do is to send a re-encryption key to the CSP and the new list of users that will have access to the selected file. The CSP will re-encrypt the file and inform the appropriate users.

If Alice wants to allow bids for part of her information, then firstly she creates the appropriate Bloom filters. Afterwards, she selects a subset of bidders from the list of certified bidders, from whom she wants to receive bids. This list contains their certificates, as in SSL, along with a public address in which they receive record IDs and decryption keys. The use of these list can lead to fine-grained policies towards accepting bids, depending on several constraints like reputation or location. So Alice creates a message that contains the record ID along with the decryption keys for the Bloom filters. Sending the keys for the decryption of Bloom filters allows her to offer a preview of her profile and avoids disclosing her data.

Having the decryption keys for the Bloom filters of some records, the Bidder sends his queries through the QR to the CSP, using the RVBM protocol. From the replies that the Bidder receives, he makes an estimation of whether he wants to access this file, the subset of files or the user's profile. This estimation is based on the number of keyword hits that he has found. So the Bidder computes a bid value. To place his bid, the Bidder uses the trading protocol described in [88] and receives a key to decrypt the data, if the user agrees.

The steps of the trading protocol are the following: The Bidder initializes a session with the CSP and the CSP creates a "ticket" for him with an "offerID" and sends it to the Bidder. The Bidder uses this ticket to contact the User and negotiates the encryption parameters. The User replies with the accepted parameters. The Bidder now is able to place his bid, which consists of the hash of the requested messages that he retrieved when he queried the CSP, the bid value or a key that releases access to a service from time t_1 till t_2 or an amount of money and the time for how long his offer will last. The user either accepts or rejects the offer and signs her reply in either case. The Bidder acknowledges user's response and if it is positive, the user will send a decryption key for the requested information to the Bidder. The Bidder can use this key to request the information from the CSP and decrypt the content that he bought. Figure 7.2 illustrates the trading protocol, while a graphical representation of the proposed scheme is shown in Figure 7.3.

A detailed analysis of the trading protocol, along with its security properties, is available in [88].

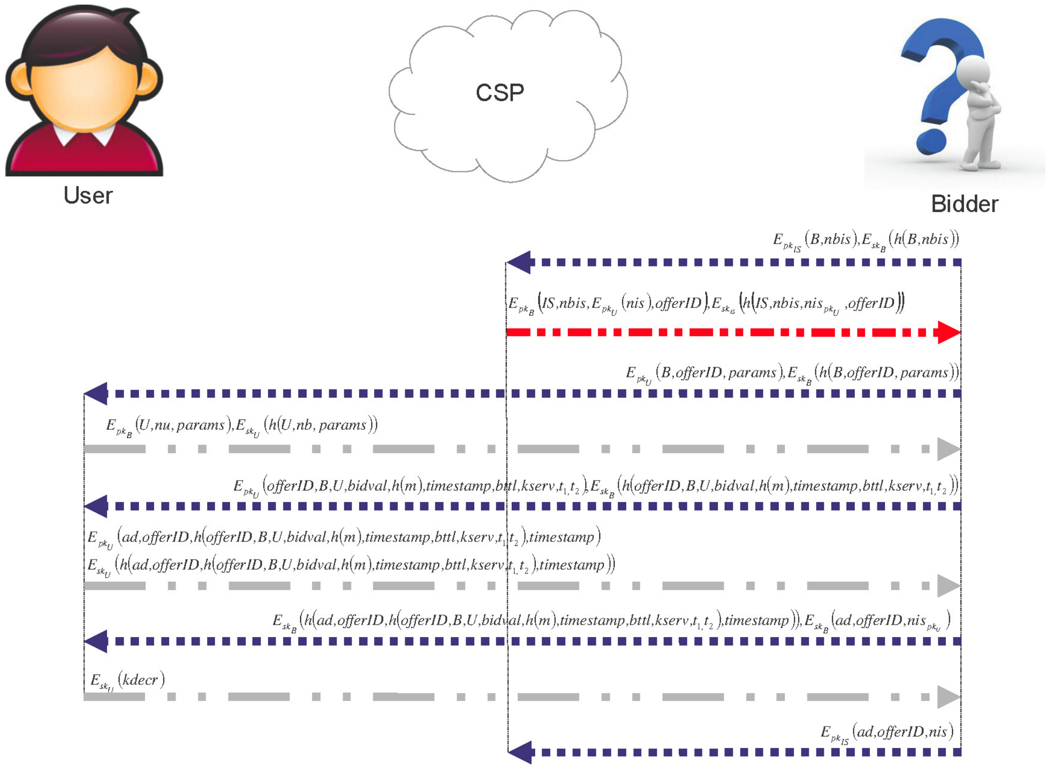
User
CSP
Bidder
$E_{pk_{IS}}(B,nbis),E_{sk_B}(h(B,nbis))$
$E_{pk_B}(IS,nbis,E_{pk_U}(nis),offerID),E_{sk_{IS}}(h(IS,nbis,nis_{pk_U},offerID))$
$E_{pk_U}(B,offerID,params),E_{sk_B}(h(B,offerID,params))$
$E_{pk_B}(U,nu,params),E_{sk_U}(h(U,nb,params))$
$E_{pk_U}(offerID,B,U,bidval,h(m),timestamp,bttl,kserv,t_1,t_2),E_{sk_B}(h(offerID,B,U,bidval,h(m),timestamp,bttl,kserv,t_1,t_2))$
$E_{pk_U}(ad,offerID,h(offerID,B,U,bidval,h(m),timestamp,bttl,kserv,t_1,t_2),timestamp)$
$E_{sk_U}(h(ad,offerID,h(offerID,B,U,bidval,h(m),timestamp,bttl,kserv,t_1,t_2),timestamp))$
$E_{sk_B}(h(ad,offerID,h(offerID,B,U,bidval,h(m),timestamp,bttl,kserv,t_1,t_2),timestamp)),E_{sk_B}(ad,offerID,nis_{pk_U})$
$E_{sk_U}(kdecr)$
$E_{pk_{IS}}(ad,offerID,nis)$

Figure 7.2: The trading protocol.

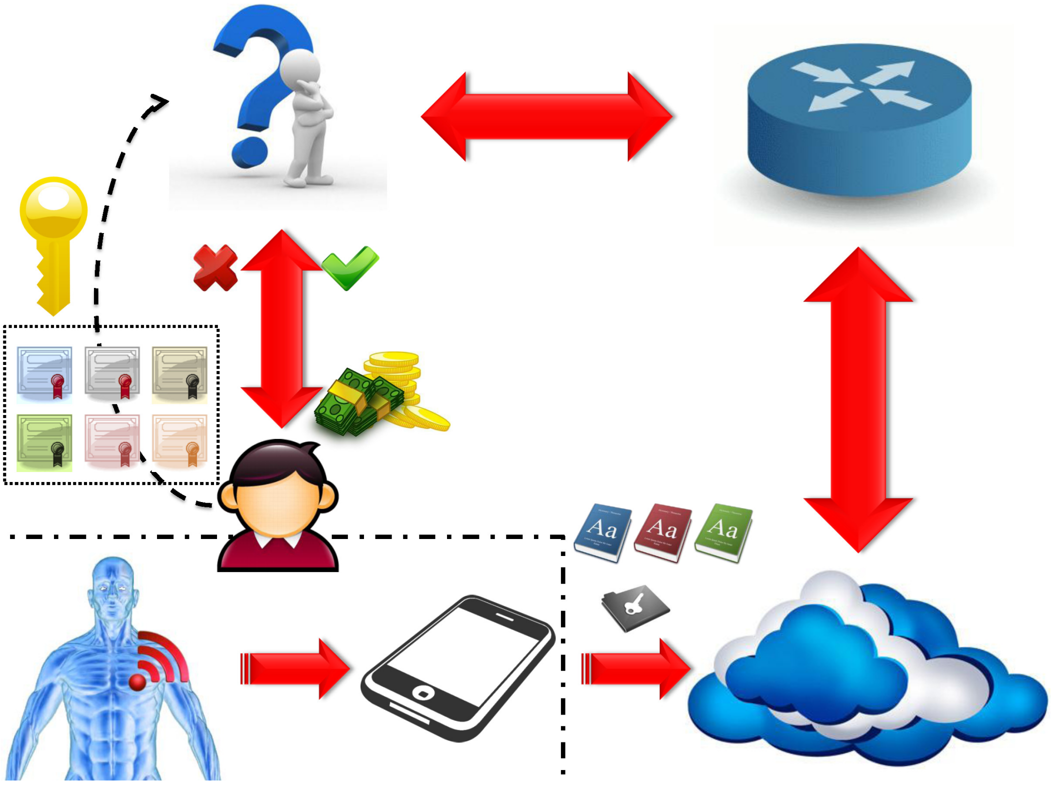

Figure 7.3: The proposed scheme for PaaP using body sensors and a Cloud Service Provider.

7.2.3 Viability of the business model

In order to prove the viability of the proposed business model, we will focus on proving the following hypotheses:

1. All parties will gain from the partnership.
2. All parties will be "forced" to play fair, otherwise they will lose more than the others.
3. The economic revenue is relevant enough to make several entities to adopt it.

From the proposed business model, all parties have something to gain:

- **Users**: The adoption of our scheme enables users to have full control over their shared information, without any unwanted disclosures. Moreover, apart from the free service, users can have additional revenues and services.
- **CSP**: The role of the CSP changes significantly from the classical model, as it is not in control of the information and it is just hosting the encrypted data. However, the CSP has other ways in order to obtain revenues. Firstly, the privacy that users gain circumvents the major obstacle that prevents many

users from using the cloud. Therefore, users will become more willing to accept and try the cloud and would even consider buying subscriptions, for example, for storage services. The CSP might not have access to the data, to trade profiles with advertising companies (third parties), however, it is the one that seals every deal and guarantees "safe product delivery". Therefore, the CSP plays a very relevant role and can claim a percentage of every deal between a Bidder and a User, or charge the QR for every executed query.

- **Query routers**: The main source of income for QRs are the bidders. QRs can charge Bidders for certification and for forwarding and anonymizing their queries.

- **Bidder**: In opposition to our business model, the current one has many limitations for the Bidders. The market is indoctrinated by the major CSPs, which charge as much as they want for their services, as there are not many alternatives. Our proposed business model manages to settle a new more solid and fairer ground. The users come in direct contact with the Bidders, so the Bidder plays a more important role in the market regulation and can minimize the prices that he pays, or even offer more services, as he has access to the source of information.

One might argue that the proposed model will work very well only if users decide to trade information, but if they adopt a selfish behaviour, meaning that they decide just to host the information and not to trade anything, then the model has nothing to offer. While claim sounds valid, nonetheless, recent figures and facts disprove it.

According to a recent study [13], the users of Dropbox, one of the major CSPs for individual users, showed this selfish behaviour. More than 95% of the users were using the "free" option, which does not, in any sense, stop them from encrypting the files, and deprives the CSP of any information. Nonetheless, Dropbox is increasing its revenues and expanding its growth. Other CSPs like SpiderOak are offering the free option for hosting information on their Cloud, providing the necessary guaranties that they have no access to the hosted information. Their revenues are only coming from paid subscriptions.

7.2.4 Fair play

The fact that each party has interest in following the protocol does not guarantee that each party will not try to bypass it in order to gain more benefits. However, next we prove that all parties have their interests towards keeping their given roles and we show it for every possible scenario of malicious cooperation. Note that in our analysis the cooperation has to do with key disclosure, since the trading protocol which seals all the deals has already proven to be secure.

The scenario that we do not consider is the cooperation between the CSP and the users, since in that case the model returns to its current form. For the same reason, the Bidder will not cooperate and disclose the purchased decryption keys to either the QR or the CSP. The only possibility is that the Bidder is closing down his business and trying to extract some last money, in every other case, he is giving back power to the CSP, which is his major interest from this business model. We have to note, however, that the CSP - QR cooperation is a calculated risk, since whenever someone decides to disclose some information, he knows that this decision has the risk of allowing third parties to access it, if the other entity is not so trustworthy.

The QR cannot cooperate with the users, so he chooses to cooperate with the CSP, then the CSP becomes aware of what queries are being made and by whom, not the results. In this case the CSP can contact other Bidders and receive the queries straight away from them, circumventing the QR. Hence from such partnership the QR only has to lose. The cooperation between the Bidders and the QRs cannot have any valuable outcome, so it is not studied.

Finally, the Bidders may cooperate with the Users. Amongst all the scenarios, this seems to be the most probable as they come in contact to exchange the decryption keys for the Bloom filters, they might as well trade the data, circumventing the CSP's part of the deal. However, it is obvious in that case the Bidder cannot be sure that he will get the requested data. If he decides to prosecute the user, he will have to publicly declare his malicious actions. It is quite clear that the same applies vice-versa.

7.3 Application scenarios

There are different scenarios where users send their data to the cloud in order to use applications and services. However, there are situations in which people want to maintain their privacy, since the shared information is sensitive. In these scenarios the scheme that we propose can be used.

A clear example of this scenario can be found in the healthcare sector, both for the general public and for professional athletes, which want to be monitored to improve their results. We analyse both cases in the next section:

7.3.1 General people

The medical records and tests that people are taking in order to monitor their health provide only a snapshot. Therefore BSN can provide a continuous mapping of patient's health, with minor physical pervasion. Currently, devices able to measure in real time the miles traveled by people, their heart rate, their movements, etc., for then analyse them on an external device and, sometimes, share them on social networks are more and more fashionable and commonplace. These data are used by the users for their own control, but they can be very valuable information to third parties also (e.g. shoe companies who want to improve their products based on the pressure exerted by the users of their shoes in their daily lives). To prevent the access to this information without users' consent, users must protect this information, which in some cases can be as critical information (e.g., Health data). Millions of people worldwide suffer from diseases that can be monitored by means of body sensors that send patients data to caregivers or a medical centres. Applications able to control the vital signs of patients and then send them to the cloud for analysis are widespread, but data can also be important for research, pharmaceutical or medical centres, which need real data to study or target advertising to a specific audience. So, it is very important that the users can protect this information and only allow its use with their consent.

Let us suppose the following scenario in which an elder person, called Abe, uses a monitoring system with body sensors and GPS. He has cognitive impairments and he is often disoriented. On one of his daily walks he goes to an ATM to get some money, but when he is coming back home he forgets that he went to the ATM and he goes again, and so on a number of times. Due to the fact that these data are stored in the cloud to be analysed, it could be intercepted by an attacker, who can deduce that Abe was in an ATM several times and that he can be disoriented and have money with him. This information could be used to perform a physical attack/robbery on Abe.

With our proposal, users are able to encrypt and split their data in the cloud, so they can use these data at same time that the data are protected. In addition, if someone is interested in buying some of these data, PaaP allows the two parties to interact in order to reach an agreement for the price of the data. To show a use case scenario of the latter form, a user named Anne, who is receiving a special medication, decides to share part of her health record, which displays how her treatment is going. The medical company that is producing the specific drugs or an competitor one want access to Anne's medical information and is willing to pay her for that in order to prove the superiority of their product, to track possible side effects, or even to fine-tune Anne's dosage to achieve better results and therefore improve their product.

7.3.2 Professional athletes

Statistics in sports is about to hit a whole new level. Due to the small size and wireless connectivity of current sensors, professional athletes can adopt these sensors without posing them a burden while training or competing. As a result, a new generation of wearable monitors that measure heartbeat rate, electrical activity in the heart, lung capacity, metabolism, and other metrics are allowing scientists to study athletes' physiology as they play. In Figure 7.4 we depict some of most common body sensors:

a) Tiny body sensors provide real-time athlete monitoring [1],

b) miCoach Speed Cell, collects speed, acceleration, distance and pace data [2],

c) Orpyx sensor embedded within the insole captures raw force data [3],

[1] http://www.gizmag.com/go/7979/
[2] http://micoach.adidas.com/
[3] http://orpyx.com/

d) The Nike+ SportWatch GPS keeps track of a user's location, pace, distance, laps, and calories burned [4],

e) Nike+ FuelBand tracks a user's daily activity including running, walking, dancing and dozens of everyday activities [5],

f) The Zephyr HxM BT is a fitness-tracking device designed for use with most smartphones. It combines Smart Fabric, heart rate sensor technology, movement sensors, and Bluetooth connectivity on a chest strap for athletes [6],

g) Special fibres bonded onto ADIDAS tops, work in conjunction with Polar's Wear Link technology to eliminate the need for a separate chest strap to monitor heart rate [7].

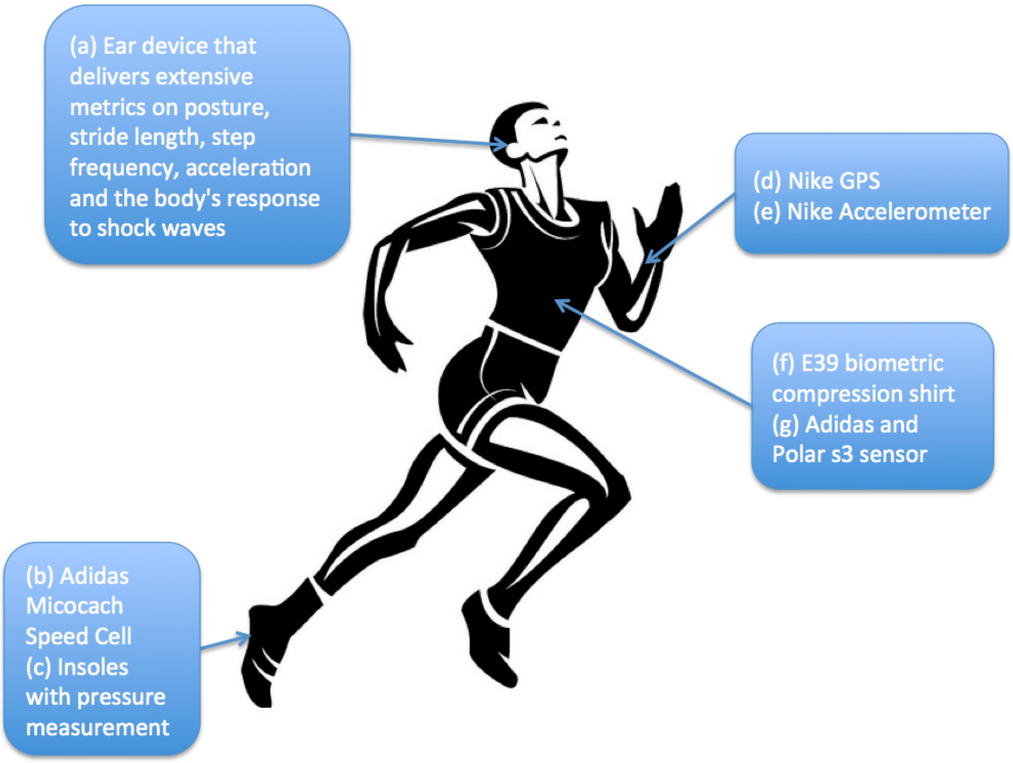
(a) Ear device that delivers extensive metrics on posture, stride length, step frequency, acceleration and the body's response to shock waves
(d) Nike GPS
(e) Nike Accelerometer
(f) E39 biometric compression shirt
(g) Adidas and Polar s3 sensor
(b) Adidas Micocach Speed Cell
(c) Insoles with pressure measurement

Figure 7.4: Some of the most common body sensors currently used by athletes.

All the data collected by these sensors have a clear potential to enhance athletes' health, and to help trainers tailor workouts. Moreover, monitoring device makers and the sports industry seems even more excited about the prospects for entertainment. They are already working on ways to display the data during games, in stadiums and on television, giving fans unprecedented insight into players game-play.

[4] http://nikeplus.nike.com/plus/products/sport_watch/
[5] http://www.nike.com/us/en_us/lp/nikeplus-fuelband
[6] http://www.zephyr-technology.com/solutions/psm-training-echo
[7] http://www.gizmag.com/go/4402/

Real-time monitoring of the wellness of athletes, during their sporting activity and training, is important in order to maximise performance during the sporting event itself and during training, as well as being important for the health of the sportsmen overall, specially for body builders, who do a lot of exercise, carrying their bodies to the limits, and they take medical drugs in order to improve their training results, but these drugs can cause serious health problems for them, so they are interested in analysing their health through the use of body sensors that monitor their data while they are training. These data are sent to the cloud to be analysed by their physicians or coaches, but doing so, if proper measures are not taken, it could endanger the athletes' privacy. Advertising companies can use their data to send them spam about new drugs, treatments, gyms, etc., without their consent.

We propose a new environment for athletic applications which, through PaaP, allows athletes to get real-time recommendations from physicians or coaches. Body builders may securely share their training results, so that they can be easily monitored by specialists.

In our proposal for this scenario the athlete hosts his/her biometric data, like heart rate, posture, acceleration, etc., on a Hosting Server (HS), splitting the data into different groups, according to their nature/type. These data are very interesting for companies in the sports sector (e.g. Nike, ADIDAS, Reebok, etc.), which could further improve their products, through crowd-sourcing. Moreover, other athletes and coaches are interested in these data, since they want to learn and be trained from the best source – the bodies of professional athletes. Therefore, athletes could be interested in selling this information, however, its sensitive nature demands a special treatment, and the PaaP framework that was previously discussed can guaranty not only privacy but a successful business model as well. Apart from sport industries, which can easily come in contact with high ranked athletes, athletes and coaches who want to perfect their skills might find this data very useful and request access to this information for some price that they negotiate using PaaP provided by the CSP.

More fine-grained privacy policies may include trajectory obfuscation and anonymization for general access and raw trajectory data for special users, using techniques as in [34, 80, 111].

Part III

Smart Cities and Applications

Chapter 8
Privacy-Aware within a Smart City

In this chapter we identify a number of privacy breaches that can appear within the context of smart cities and their services. We define the concept of Citizens Privacy as a model with five dimensions, namely Identity Privacy, Query Privacy, Location Privacy, Footprint Privacy and Owner Privacy. By means of several examples of smart city services, we define each privacy dimension and we show how existing Privacy Enhancing Technologies could be used to preserve Citizens Privacy.

Contents

8.1 Introduction

Countries struggle to be competitive, attract investments and talent, reduce debt and be globally sustainable. Due to factors related to economies of scale, many services are more easily provided in highly populated areas. Hence, people are moving from the country to the cities and a urbanization trend starts to be apparent throughout the world. As a result of this urbanization process, cities are gaining importance and their role as economic engines is becoming more prominent nationally and also at an international level.

The struggling of countries for competitiveness has a smaller counterpart in the shape of their cities. Those cities are internationally competing for investments, talent and even to increase tourism, and they realize that the most promising path to success requires the use of technology. Thanks to Information and Communication Technologies (ICT), local governments and private companies like Cisco and IBM are developing and implementing innovative solutions to improve the management of cities operations in a variety of areas, namely transportation, energy, sustainability, e-governance, economy, communications, and so on.

Although the concept of *smart city* is pretty new, we can find several examples of cities that have adopted it. For example, the city of Amsterdam [70] has defined four areas (*i.e.* sustainable living, sustainable working, sustainable mobility, and sustainable public space) around the idea of sustainability, in which smart projects are conducted so as to improve the city and transform it into a real smart city in the near future. In Amsterdam, they focus on the reduction of CO_2 emissions but there are other approaches focused on reducing the cost of public services and transportation [61], improving the interaction of the society with the administration, or simply improving the experience of tourists. Some other examples of cities working towards the "smart" line are Vienna, Toronto, Paris, New York, London, Tokyo, Copenhagen, Hong Kong and Barcelona [30].

The fundamental rights of citizens should be guaranteed anytime. In this regard, for smart cities to be a successful reality, we emphasize the importance of the preservation of privacy. Most of the services offered in smart cities are based on ICT. Users interact with these services through a wealth of devices (*e.g.* smartphones, information totems, public computers, etc.) that are connected using heterogeneous networks and systems, which are the perfect target for attackers and eavesdroppers willing to disclose sensitive information from individuals or even to impersonate them. In addition, the huge amount of data collected and managed paves the way to the Big Brother effect. As a result, citizens might be refrained from using the smart city services to avert such problems.

Legislation is essential to guarantee the achievement of privacy within smart cities. Individuals must be aware of the ability of smart cities to silently gathering a variety of information about them. Hence, the wide adoption of legislation regarding the collection and processing of personal data [31] within a smart city would be the icing on the cake.

Last but not least, although technical solutions (e.g. encryption, digital signa-

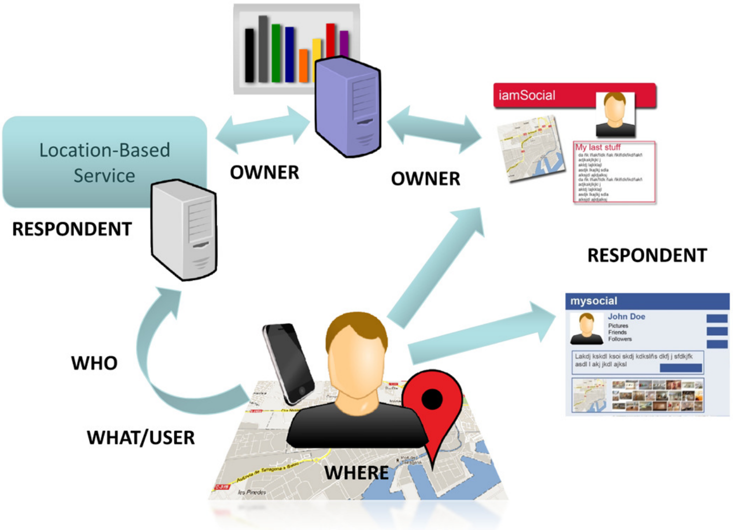

Figure 8.1: Conceptual scheme of the privacy models. In the picture we distinguish (i) a user (at the bottom) that contacts a location-based service provider, (ii) two social networks (on the right) to which the user belongs, and (iii) a data warehousing facility (at the top) [73].

tures, server reliability, etc.) make smart city services feasible from a security point of view, there is still a lot of work to be done so as to materialize the notion of privacy in smart cities.

We present the concept of *Citizens Privacy*, which consists in the application of the so-called Privacy Enhancing Technologies (PET) in the smart city scenario. We show that a combination of these techniques –currently used in privacy models for databases and location-based services–, can be applied to build a model for Citizens Privacy.

8.2 Privacy Models

In this section, we recall two privacy models that can be applied to achieve Citizens Privacy: the 3-Dimension Conceptual Framework for Database Privacy [35] and the W^3−(Where, Who, What) privacy model for location-based services (LBS) [91].

Figure 8.1 illustrates the theoretical privacy dimensions described in this section.

8.2.1 The 3D Conceptual Framework for Database Privacy

An astonishing amount of data from multiple sources is collected and stored in databases belonging to multiple parties (*i.e.* governments, private companies, etc.). The privacy of the data stored in these databases might be understood differently depending on the context and the operations applied. Domingo-Ferrer [35] splits database privacy issues into three dimensions related to the main actors involved, namely respondents, users and owners:

- **Respondent privacy.** It is focused on avoiding the re-identification of individuals (*i.e.* respondents) whose information is stored in a database. In the example of Figure 8.1, the user queries an LBS provider and publishes his activities in social networks. These data are stored in the databases of the service providers and can be analyzed to obtain a variety of information. Regarding respondent privacy, no sensitive or private information should be leaked from these databases. They must be protected before being published or released to third parties. *Statistical Disclosure Control* (SDC) is usually used to do so.

- **User privacy.** This is about guaranteeing the privacy of the queries made by a user to a database system (*e.g.* Internet search engines, LBS providers). The point is to obtain the desired information without revealing the real query to the database system. This is known as the *Private Information Retrieval* (PIR) problem. In relation to our example, the queries made by the user to the LBS provider should follow a protocol to prevent the provider from learning them.

- **Owner privacy.** This privacy dimension refers to the owner of a database queried by other users/entities. The owner might agree to share some of his data but it should be controlled that only those data (and no more) are gathered by the issuers of the queries. The *Privacy Preserving Data Mining* (PPDM) discipline designs techniques to address this problem. In our example, a third party (a data warehousing facility) pays an LBS provider and a social network to mine their data. In this case protecting owner privacy means to allow the third party to access the information he paid for but no more.

The aforementioned techniques (SDC, PIR and PPDM) are described in Section 8.3 below.

8.2.2 W^3-Privacy for Location-Based Services

As we stated in previous chapters, services related to the location of the user are gaining importance and so do privacy issues related to them. In 5 we describe the three dimensions of user privacy in LBS and define the concept of W^3-privacy. Those dimensions can be inferred from the main parts of a typical location-based query: "**Someone** is asking for **something** near **somewhere**":

- **Where.** This is the privacy dimension related to the location of the user. LBS providers might learn that location from the queries of the user. Thus, users could be tracked. In our example, the user sends his current location to the LBS provider to obtain an answer. Thus, the LBS provider may track him. Several techniques have been proposed to mitigate this problem (*e.g.* collaborative location obfuscation, cloaking, etc.).

- **What.** In general, LBS providers inform users about *something*. The *What* dimension of privacy in LBS refers to the privacy of the queries. Note that this dimension is very similar to the *User Privacy* dimension in the databases context. Hence, PIR techniques can also be used to approach it.

- **Who.** This problem is about identifying the user and relating him with a bunch of queries. This might allow the provider to create user profiles. In order to mitigate this privacy issue, most solutions rely on intermediate entities to hide real identities using, for example, temporal pseudonyms.

In Chapter 5 we delved into the W^3-Privacy concept, where an LBS is W^3-private if the service is given while the LBS provider cannot know: (i) who is the user, (ii) where is the user, and (iii) what does the user ask for.

8.3 Privacy Enhancing Techniques for Citizens Privacy

In this section, we describe the techniques that can be used in our Citizens Privacy model, namely Statistical Disclosure Control, Private Information Retrieval, Privacy-Preserving Data Mining, Location Privacy, Anonymity and Pseudonyms, Privacy in RFID and Privacy in Video Surveillance. Their use will be illustrated in the example addressed in Section 8.4.

8.3.1 Statistical Disclosure Control (SDC)

Private companies and statistical agencies collect data from people in a daily basis. On the one hand it is necessary to guarantee the right of the society to information but, on the other hand, the right to individual privacy should be preserved. The field of Statistical Disclosure Control aims to protect the privacy of individual respondents while allowing the release of their data for secondary use. Many techniques have been proposed to protect respondents privacy, namely noise addition, microaggregation, rank swapping, rounding, etc [60]. The main aim of these techniques is to distort data so as to avoid the linkage of private information with individual respondents. At the same time, the distortion introduced into the data should be limited to preserve data utility. All in all, statistical disclosure control techniques try to find the right balance between information loss and disclosure risk. These topics are addressed and formalized under the methodologies of *Differential Privacy* [37].

8.3.2 Private Information Retrieval (PIR)

Consider the problem in which a party A wants to obtain a piece of information from a database belonging to another party B. A wants that information but it does not want B to know which it is. This problem is known as the Private Information Retrieval problem and Chor, Goldreich, Kushilevitz and Sudan introduced it in 1995 [27]. The simplest solution for A to achieve its goal is to ask B for the whole database. If B sends the database to A, it is impossible for B to know which is the information that A is interested in. However, this trivial solution is not practical due to communication costs. Since the problem was stated in 1995 a number of protocols have been proposed to reduce the computational and communication costs [120]. However, in general PIR approaches are considered to be impractical in real scenarios yet.

8.3.3 Privacy-Preserving Data Mining (PPDM)

Due to the ability of ICT for gathering unprecedented amounts of data, data mining techniques gained much attention. The main goal of data mining is to develop models representing aggregated data so as to discover non-obvious, valuable data. More generally, we might say that data mining aims to obtain knowledge from data. However, due to numerous privacy concerns, data mining was seen as a privacy threat and the field of Privacy-Preserving Data Mining appeared to change data mining for the better, providing all its benefits while maintaining privacy [115]. In general the PPDM problem can be seen as a game between two parties that do not trust each other. Both parties have some data and need to collaborate to obtain a common result but they do not want to share their data. Many protocols have been proposed to approach this problem from simple data perturbation techniques to the more sophisticated multiparty computation.

8.3.4 Location Privacy

When users try to obtain information from an LBS provider, they send their location and allow the LBS provider to track them. Several methods have been proposed to protect location privacy. Their aim is to provide a distorted location that prevents the provider from tracking users. In [43] the authors propose the use of a trusted third party (TTP), which handles users locations to create cloaking regions. Users send these regions to the LBS and, since several users are under the same cloaking area, the server will not be able to correlate users and locations. Other proposals that do not rely on TTP also exist but, require several protocol rounds and/or users collaboration [105].

8.3.5 Anonymity and Pseudonyms

When users contact a service to obtain information, their identity is exposed to the provider and it can link users with their queries (which might lead to profiling and thus, invasion of privacy). To address this issue, most solutions rely on intermediate entities to hide the real identities of the users (*e.g.* using pseudonyms). Also, TTP-free versions based on collaboration among users have been proposed [94].

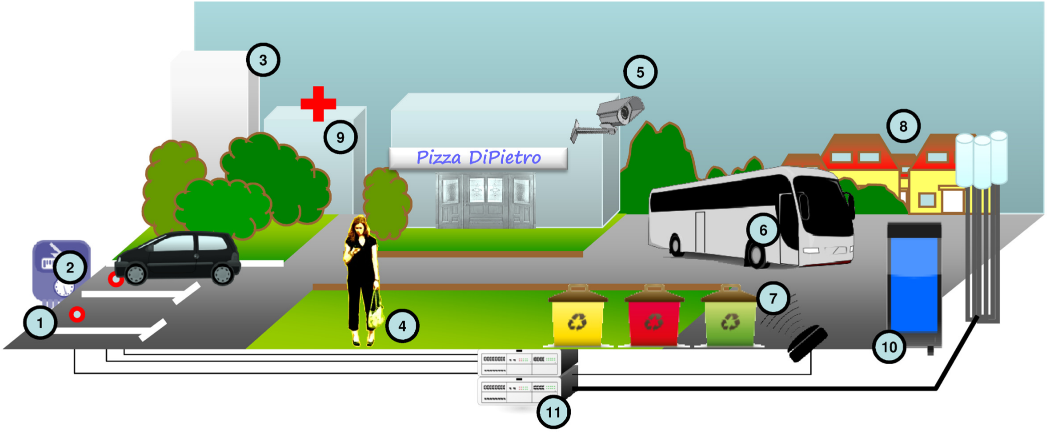

Figure 8.2: Our example of smart city: (1) Smart parking service (2) Electric car recharging (3) Smart building with control of presence (4) User querying an LBS provider (5) Camera for video surveillance (6) Smart bus that changes the route upon users needs (7) Smart garbage containers (8) Control of energy consumption in homes (9) Medical center (10) Interactive information pole (11) Network infrastructure of the smart city [73].

8.3.6 Privacy in RFID

Radio Frequency IDentification (RFID) systems consist of tags and readers. Tags contain identification information of products that can be accessed by readers without the need for visual contact and cables. This is very convenient for the manufacturing sector but might be a privacy problem if unauthorized people could read tags and obtain their confidential information (and by extension the information of the user). With the aim to solve this problem many protocols have been proposed and it could be said that privacy and security can be guaranteed. However, the main problem is to achieve privacy and security in reasonable times (*i.e.* there are scalability problems [108]).

8.3.7 Privacy in Video Surveillance

Pervasive video surveillance systems inherently endanger the privacy of people: identities and activities can be easily retrieved from pictures and videos. People accept to be controlled for the sake of security, but most privacy advocates warn about the Big Brother effect. In [74], the authors claim that video surveillance systems must guarantee the private management of video data. To that end, they use real time computer vision techniques to accurately detect regions of interest (*i.e.* faces, car plates, etc.), that are then protected.

8.4 A 5D Model for Privacy in Smart Cities

Researchers, practitioners and administrators must take into account the privacy concerns that entail the pervasive nature of ICT in smart cities. To that end, we propose the concept of Citizen Privacy: a 5-dimensional model for Citizens Privacy in smart cities. The identified dimensions are: *Identity* privacy, *Query* privacy, *Location* privacy, *Footprint* privacy, and *Owner* privacy.

Next, we define each dimension in the context of a smart city. For each dimension, we show examples of privacy concerns. Also, we point to the technologies (introduced in Section 8.3) that could be used to address those concerns. The examples used throughout the section refer to Figure 8.2. The goal of the scenario depicted in Figure 8.2 is to provide a non-exhaustive but illustrative set of real smart city services. In the figure, we illustrate the following services:

1. *Smart parking service.* In this service, available parking spaces are controlled by sensors. Drivers are guided to the nearest available parking space, and they pay for the exact time that they use the parking space.

2. *Electric car recharging.* Complementing the parking service, electric cars can use recharging sockets. Users pay for the energy they consume to charge their cars.

3. *Smart office building.* This service controls who is in their office in order to optimize the energy consumption related to illumination, air conditioning, etc.

4. *Location-based service.* This service allows the query of information based on the location of the requester. In our example, a citizen is looking for a list of nearby Italian restaurants.

5. *Video surveillance system.* For the sake of citizens safety, the city is covered by a network of pervasive and interconnected cameras.

6. *Smart bus service.* This bus optimizes its route in real time according to the number of users that request its service.

7. *Smart garbage containers.* These containers send an alarm when they are full and need to be emptied. Moreover, only users living in the surrounding area can use them.

8. *Control of power consumption at homes.* In order to improve the production and distribution efficiency of energy, the consumption levels are collected via a sensors network.

9. *Medical center.* The medical center collects data from patients. Moreover, personnel in the medical center query other hospitals to retrieve information about the patients they are in charge of.

10. *Interactive information pole.* Users can access these devices to obtain information about the city. In addition, citizens identify themselves and access personalized services.

Our citizen privacy model, including examples and solutions, is summarized in Table 8.1.

8.4.1 Identity Privacy

Definition. Identity privacy relates to disclosing the identity every time a user accesses a smart city service. In that sense, it is mapped to the *Who* privacy of the W^3-privacy model. If users specify their identity, service providers and other third parties will be able to correlate users and their activities.

Examples. This is a common issue in many services. Users disclose their identity when they access the smart parking service, or when they pay for the car energy recharging service. Moreover, the detection of occupancy in the areas of the smart building could also entail identification. Also, as we have pointed out previously, the use of LBS generally entails identifying the user. Last but not least, the video surveillance system clearly involves identity privacy concerns.

Solution. The use of pseudonymizers contributes to preserve identity privacy. One could think of a single pseudonymizer service. However, if the service is attacked or their administrators misbehave, the relation between identities and pseudonyms can be disclosed. To avoid that situation, this service should be provided by a set of geographically distributed pseudonymizers. Finally, with regard to video surveillance, real-time, accurate protection of the regions of interest might be applied.

8.4.2 Query Privacy

Definition. Query privacy is related to preserving the privacy of the queries made by users to services. Hence, it is mapped to both *User* and *What* privacy dimensions. Upon collecting the queries made by users, service providers can profile users and obtain information about their habits.

Example. The interactive pole and the LBS involve this privacy issue. Moreover, services such as smart parking and smart bus may also entail query privacy, since the queries made by users can be analyzed to extract information about habits.

Solution. In general, PIR-like techniques might be used to mitigate the query privacy issue: services should include PIR tools that users might apply before querying the provider. Using TTP might also be an option. Whatever technique is applied, the goal should be hampering the correlation of users and queries.

8.4.3 Location Privacy

Definition. Location privacy is about guaranteeing that the privacy of the physical location of the user is preserved. This is the *Where* dimension of the W^3-privacy model.

Table 8.1: Summary of our 5D proposal for modeling the privacy aspects in smart city services [73].

5D approach	**Existing models**		**Examples of privacy concerns**	**Existing solutions**
	3D Database	**W3 LBS**		
Identity		Who	Most of the examples entail identity privacy concerns. RFID and video-surveillance are also related to identity issues.	Pseudonymizers, RFID privacy techniques, privacy-aware video surveillance
Query	User	What	Mainly location-based services, interactive information poles, etc.	Private Information Retrieval techniques, random pseudonymizers.
Location		Where	Location-based services, other services involving location (for example, smart parking). Also video surveillance entails location privacy.	Collaboration for location masking, cloaking, pseudonymization, privacy-aware video surveillance.
Footprint	Respondent		Microdata generated from a variety of sources (sensors, RFID readers, medical data, electronic voting, etc.)	Anonymization, Statistical Disclosure Control
Owner	Owner		Obtaining information across databases belonging to different entities.	Privacy-Preserving Data Mining, Statistical Disclosure Control

Example. Clearly, the LBS of our scenario entails location privacy issues. However, almost all the depicted services also entail them: using the smart parking service, users disclose their location in order to be routed to the nearest parking area; using the car recharging service, the location of user's car is disclosed; the smart building is also aware of the location of individuals; etc.

Solution. In some cases in which the location is not constant, LBS users could collaborate to mask their exact locations. Also, a cloaking service could be used to protect real locations.

8.4.4 Footprint Privacy

Definition. Footprint privacy is related to the control of the information that can be retrieved or inferred from microdata sets. Actually, the activities in a smart city involve the acquisition, collection and storage of large amounts of microdata *i.e.* the information at the level of respondents. In our definition of footprint privacy, these microdata are obtained from a variety of sources, namely sensor networks, RFID readers, etc. Hence, a service is related to a microdata set that records the information about the use of the service (that is, the *footprint* of the users on the service). The microdata sets can be published or released to third parties so the latter can obtain a variety of information. The privacy of the individuals must be preserved and, hence, the disclosure of sensitive information should not be possible from the released data. Therefore, this dimension can be mapped to the *Respondent* dimension that has been described for database privacy.

Example. All the services that involve the acquisition of information about their utilization may suffer from footprint privacy issues.

Solution. The aforementioned SDC techniques must be applied over the data sets before their release. Besides deleting the identification information (or at least replacing them by pseudonyms), some procedures should be performed to control the disclosure risk while restraining the information loss.

8.4.5 Owner Privacy

Definition. Owner privacy deals with the privacy-aware computation of queries across the databases from different autonomous entities. This dimension is directly borrowed from the *Owner* dimension described for database privacy.

Example. Let us focus on the energy consumption control in homes and assume that the electricity company wants to correlate the use of electricity with the use of other services such as telephony or gas. A naive solution would be that the telecommunications and gas companies released their *footprint* databases to the electricity company. Naturally, the knowledge extracted from these databases is highly attractive for strategic and commercial decisions and, consequently, these companies may refrain from releasing or sharing their data.

Solution. The owner privacy issues are the natural scenario for PPDM techniques and even SDC. If they are applied to the queries across the databases, the amount of information actually transferred to the entity that originates the query will be controlled.

8.4.6 Towards Smart Cities

Although economies of scale help to reduce costs, managing cities is difficult because the number of inhabitants is growing steadily and the infrastructures and procedures have to be adapted to a growing and very demanding population. Thus, local administrations have the need for smart procedures to improve the quality of life and the management of resources in their cities.

Many cities are starting to use technology to improve the quality of life of its inhabitants. Examples of technology that they are using are: sensors to control lighting when there are people around, location-based services, smart parking or surveillance cameras. This technology helps to reduce costs while improving its infrastructure to provide better quality of life. But creating these intelligent environments introduces new risks to the citizens' privacy and that raises some questions as "Is it worth to scarify your privacy in exchange for a better quality of life?".

Privacy protection should be considered as a real issue to face before deploying new technologies in cities, as citizens rights must me over the interests of the cities and governments. In order to face the dangers introduced by a Smart City we first need to identify and understand them. In this chapter we have identified and described some of the main privacy risks within a smart city and we have proposed the deployment of existing protocols to face each of these risks. Thus, we want to make people to understand that the introduction of new technologies in cities can live in harmony with the privacy of its inhabitants, improving their quality of life without scarifying their rights.

Chapter 9
Private Monitoring of Elderly

In this chapter we present the concept of m-Carer *as a smart mobile device able to privately monitor the movements of patients having diverse degrees of mobility and autonomy. After justifying the need for privacy-aware m-carers due to social and economical reasons, we propose a complete architecture aimed at fulfilling the needs of patients, relatives and healthcare services. Moreover, we show a real implementation of our proposal so as to confirm that it is technically sound and feasible.*

Contents

9.1 Introduction

In the last decades the society of the so-called developed countries has changed significantly. Whilst fertility rates are reaching unprecedented low figures, life expectancy grows steadily. As a result, we are witnessing the dawn of an aged society that poses new social and economical challenges. This ageing of the population leads to an increase in the cases of cognitive disorders related to age like Mild Cognitive Impairment (MCI), Frontotemporal Dementia, Lewy Body Dementia, Parkinson's Disease and Alzheimer's Disease (AD).

We are specially interested in MCI because it can be seen as a precursor of the early stages of AD and other types of dementia that imply impaired memory function whilst the cognitive function is generally preserved [95]. People suffering from MCI and early stages of different types of dementia might experience a decrease in their cognitive capabilities but they still have considerably high degrees of autonomy (*i.e.* they can live alone, walk, do exercise). The most apparent impairment is related to their memory function: patients might become spatially and temporally disoriented, and might have problems in finding their way home.[1] Note that although patients with MCI and early stages of dementia are the ones that motivated this research, patients suffering from more advanced stages of dementia might benefit even more than the formers because they tend to become spatially and temporally disoriented more frequently.

With the promise of being helpful to address many medical situations, information and communication technologies (ICT) have attracted the attention of the medical community, from physicians and health scientists, to regulators and governments. In this regard, a collection of devices and complex systems such as computers, sensors and databases are used in the so-called electronic health (**e-health**). ICT might be used for a variety of health-related tasks, namely communication between patients, doctors and carers (mobile health support), remote provision of care (mobile telemedicine), remote support to diagnostic (mobile telediagnosis), electronic medical records, smartcard-based prescriptions, etc.

The use of ICT in the healthcare sector has significantly contributed to reduction of management costs and efficiency increase increase. In this line, e-health substantially reduces the displacements of professionals and patients, globally brings down the cost of medical resources, and makes treatments and health watchfulness more comfortable to patients. All in all, e-health might be considered a revolution in this area. However, the next and probably more important revolution is taking place due to the use of mobile devices (*e.g.* smartphones) with unprecedented processing and communication capabilities. These mobile devices have favoured the emergence of mobile health (**m-health**), understood as the discipline founded on the use of mobile communication devices in medicine, or more specifically the delivery of healthcare services via mobile communication devices.

The use of mobile devices simplifies and makes many tasks more efficient. Specially the remote monitoring of patients and the communication between professionals, relatives and patients will highly benefit from the use of m-health. Moreover, m-health allows rapid gathering of data from patients, thus, providing doctors and scientists with a large amount of information that can be used for a variety of purposes.

Certainly, m-health fosters innovation in the healthcare system of industrialised countries. Nonetheless, it can be regarded as a way of providing quick and effective access to health services to large populations without nearby hospitals or medical centres. Hence, the rapid proliferation of affordable smartphones in rural areas or in developing countries [116] might clearly improve the overall efficiency and coverage

[1] When the disease advances these patients may become wanderers and might require special supervision.

of their healthcare systems. Despite all the advantages of e-health and m-health applications, they cannot be applied without considering fundamental issues like the privacy and security of the users [77] (*e.g.* data access control, data disclosure, data privacy, etc).

There is a real need for systems and methods allowing the private supervision of patients, specially when, due to their disorders, they can easily get lost. Under specified circumstances (*e.g.* when a fall is detected), these systems should permit authorised users to locate patients. Notwithstanding, if no special events happen, the system must prevent users from locating patients, thus, preserving their fundamental right to privacy.

9.1.1 e-Health concerns

e-Health has demonstrated to improve many healthcare-related procedures and we believe that, thanks to the steady advances in mobile communications, m-health will lead to the next revolution in the healthcare sector. In [107], authors introduce the new concept of smart health, and they provide an overview of the main fields of knowledge that are involved, which is the context-aware complement of mobile health within smart cities.

In this chapter we propose the new concept of **m-Carer**, this is, an intelligent application that runs on smartphones and allows the private monitoring of people's location. By using the proper cryptographic primitives, our proposal guarantees the privacy of patients whilst, at the same time, they can be located if necessary (*e.g.* if an emergency arises). Our proposal is easy to use and it does not require patients to use annoying and indiscreet devices like necklaces or bracelets.

Although we focus on patients diagnosed with MCI and initial stages of dementia, our solution could be easily extended to deal with the monitoring of children, disabled people, etc. We realise that privacy is not always a priority, specially when a patient's well-being might depend on quick and easy access to information [57]. However, we deem confidentiality of information to be critical both to guarantee the fundamental right of individuals to privacy and to avert hindrance to the monitoring system.

With the m-Carer we do not aim at replacing human carers but to provide them with a powerful tool able to simplify their job, improve their efficiency, reduce costs, and keep the fundamental rights of patients fully guaranteed.

9.2 The m-Carer architecture

In this section we provide an overview of the concept of m-Carer and its fundamental architecture. First, we justify the need for this kind of system and explain its main functions in Section 9.2.1. Second, in Section 9.2.2 we present a running example that is used throughout the chapter to support the explanation of different situations and concepts. Finally, in Section 9.2.3 we describe the main actors of the system, their roles, and their relations.

9.2.1 Rationale, concept and desiderata

As we have previously stated in the introduction, we observe two very important trends: (i) our society is getting older, thus increasing the number of cases of MCI and dementia, and (ii) mobile communications are experiencing a massive development that leads to the appearance of m-health.

m-Health redefines the healthcare services in three main aspects:

- m-Health allows **easy access** to an unprecedented number of services and knowledge. Thanks to the inherent **ubiquity** of mobile devices, services may be accessed everywhere, anytime. Moreover, data could be collected more easily regardless of the location of the user.
- m-Health is **user-oriented**. Users play a key role in an m-health service. Services should be where the user is.
- m-Health is **personalised**. Users receive customised services, that fit their needs properly.

With all these main trends and features in mind, we can define an m-Carer as follows:

> *A mobile carer **m-Carer** is a mobile device and an infrastructure that provides patients/users with healthcare monitoring services in a private, reliable, and personalised way.*

An m-Carer (in the context of an ageing society) is aimed at improving the quality of life of people suffering from MCI and early stages of dementia. To do so, an m-Carer privately monitors the location of patients and allows their safe recovery if they get lost or disoriented[2]. Also, the quality of life of their relatives and carers might be improved. Note that patients with MCI and initial stages of dementia have a significant degree of autonomy. In this regard, an m-Carer is designed as a non-invasive tool aimed at helping relatives and human carers to supervise the regular activities of patients whilst respecting their fundamental right to privacy.

An m-Carer should have the following features:

- **100% private:** The supervision should be private to avert any hindrance from the patient. Avoiding the "Big Brother" effect is paramount to guarantee the acceptance of patients with a high degree of autonomy. If alarm conditions are not met the location of patients should be kept secret.
- **Intelligent and reactive:** The m-Carer must be able to detect situations in which the safety of the patient is in danger. Some examples of those situations are the following:

[2] Note that patients with MCI and early stages of dementia are autonomous but might get lost, thus endangering their safety.

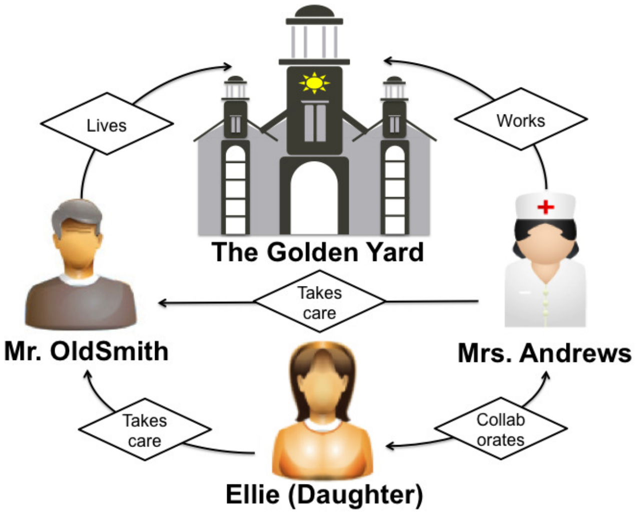

Figure 9.1: Illustration of our running example. In the figure, the main actors of the systems are depicted along with their relations [106].

  - The patient is approaching risky areas (*e.g.* highways, cliffs, rivers, etc).
  - The patient is not in a usual place and he or she is probably lost.
  - The patient has fallen and might be injured.

- **Autonomous:** Collaboration of the patient should not be required. The m-Carer might respond by its own to the possible situations in which the safety of the patient could be in danger. In those cases, the m-Carer should automatically warn the patient, their relatives or carers, and even the emergency services (if necessary).

- **Easy-to-use and accessible:** The m-Carer must be simple and easy to use (even to those patients that are not familiar with information and communication technologies). Alarms and warnings should be easy to configure by patients, relatives or human carers by means of off-the-shelf technology (*e.g.* a regular website). Also, if there is an emergency, the location of the patient might be easily accessible (*e.g.* through SMS messages, or using a secure website).

- **Discreet:** The embodiment of the m-Carer should be simple and usual. If so, patients will have no reason to reject it. It should not be seen as something strange nor intrusive.

9.2.2 Running example

Throughout the chapter we will refer to the following example so as to show the specific functioning of our proposal. The main aim of this fiction is to help us show that

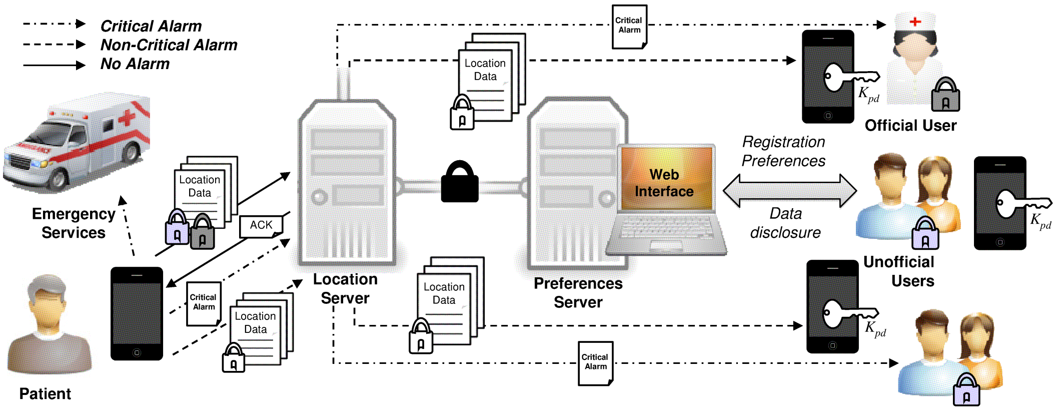

Figure 9.2: General scheme of the m-Carer system with all the actors (*i.e.* the m-Carer, the patient, the official and unofficial users, the servers and the emergency services), and their relation under different alarm states [106].

although the architecture is quite abstract it has a clear and direct application to real life situations. Note that any resemblance to real people is purely coincidental.

John Oldsmith' case: *John Oldsmith (70) was diagnosed with Alzheimer disease a year ago. All began by forgetting simple things, such as the name of his daughter Ellie or the name of the company in which he had worked for 35 years. Being a widower for three years, Mr Oldsmith decided to retire in a renowned old people's house – The Golden Yard. He wanted to live his life autonomously, together with some of his best friends who also had chosen to retire in The Golden Yard. Having large promenades and enjoying interesting after-meal talks became his everyday activities.*

Mr. Oldsmith is aware that his cognitive capabilities are going to decrease with each passing day. However, he feels strong and healthy enough to carry on with his daily walks. Notwithstanding, Mrs. Andrews, the head of The Golden Yard carers, asks John to carry a GPS-enabled necklace that allows her to know the location of Mr. Oldsmith at any moment – "just in case", she always says. Mr. Oldsmith feels that his privacy is being invaded, and he has to choose between his promenades and his privacy.

9.2.3 Structure, actors and roles

The m-Carer (an essential actor of the system) is an application that runs on a smartphone able to locate itself by means of GPS, WiFi, or fixed antennas trilateration. It is able to encrypt information and to send it to a server or a set of servers that store it. The m-Carer uses the telecommunications infrastructure that already exists and does not require any additional device (apart from the smartphone).

In addition to the smartphone (with the m-Carer) and the servers, the system considers a number of human actors, namely patients, official users and unofficial users. Figure 9.2 depicts a general scheme of the structure of the m-Carer system with all the actors and their relations.

Next, we describe those actors in detail and the role they play within the whole logical structure of the m-Carer system.

Patients

Patients are people that are able to move autonomously (*e.g.* have a walk, go shopping, etc.) that due to their cognitive impairments might get lost and require assistance. In order to guarantee their fundamental right to privacy and to avert the "Big Brother" effect, the m-Carer encrypts their information. Patients are equipped with a GPS-enabled smartphone connected to the Internet. We refer to this device as *Patient's Device* in which the m-Carer application runs.

In our example *John Oldsmith* is the patient. Mr. Oldsmith owns a simple smartphone, equipped with GPS and a 3G data connection. He has installed our m-Carer application in his smartphone and he has signed in the system.

Users

Users are people ethically or legally responsible for the supervision of patients (*e.g.* human carers, relatives, friends, etc). Although the privacy of the patients is one of our main priorities, for the sake of safety, users are allowed to know the location of patients in some specific situations[3]. Similarly to patients, users are equipped with a smartphone running a *User application.* By using this device, users can interact with the system and receive notifications. We distinguish two different types of users:

- **Official users** are people that are legally responsible for the well-being and safety of patients (*e.g.* civil servants such as social workers, doctors, nurses, or public human carers).
  In our example, *Mrs. Andrews* is the official user for all the patients living in The Golden Yard.

- **Unofficial users** are other people related to the patient, namely relatives, friends and private human carers. Those people might be in charge of patients and would need to know where they are, specially if they tend to get lost.
  In our example, *Ellie* (Mr. Oldsmith's daughter), is the unofficial user responsible for the safety of Mr. Oldsmith.

Emergency services

These are actors that ensure public safety by addressing emergencies (*e.g.* firemen and rescue services, medical emergencies, etc.). In normal conditions, these emergency services do not play an active role in the system (*i.e.* they do not directly

[3] In the next section we will discuss the details about the operation of the system and which procedures the users have to follow in order to obtain the location of patients.

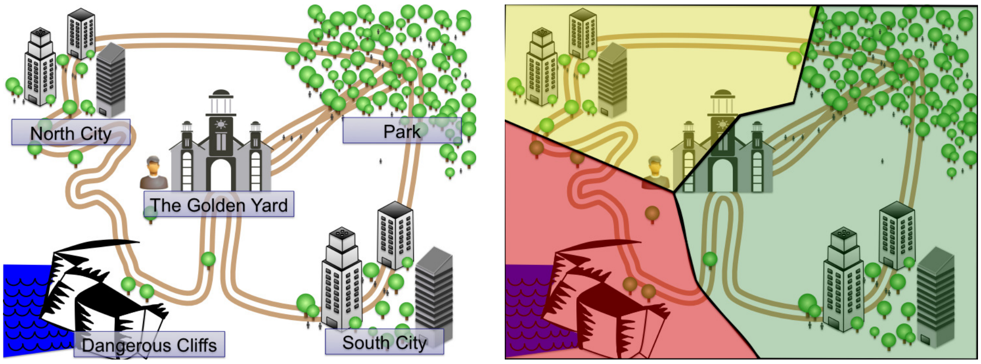

Figure 9.3: **Left:** Illustration of the area surrounding The Golden Yard in our running example. **Right:** Areas defined by patients or users (in this case *Mr. Oldsmith* and his daughter *Ellie*). Green areas indicate allowed zones without risk, yellow areas indicate non-allowed zones (possibly risky), and red areas indicate dangerous zones that must be avoided [106].

interact with the system). However, their help is required when an emergency situation arises. The m-Carer is responsible for informing emergency services and for asking for their help when a critical emergency is detected. We give more details about this procedure in the next section.

Servers of the system

In addition to the applications that run over smartphones (*i.e.* the *m-Carer* in the patient's device, and the *User application* in the device of the users), the system comprises a set of servers that support its operation and provide additional functions. Mobile applications interact with these servers by using the already present infrastructure of the telephony company and the Internet service provider. We distinguish two different types of servers in the system:

- **Servers of preferences**, that allow the registration of patients and users into the system. They are responsible for the generation of the cryptographic keys to be used by the *m-Carer* and the *User applications*. Also, they allow users and patients to tune several parameters of the system that controls its behaviour. By doing so, users and patients can customise the service provided by the m-Carer and make it to better suit their needs. Figure 9.3 shows an illustration of the surroundings of The Golden Yard in our example and the customised areas defined by *Mr. Oldsmith* and his daughter *Ellie*. The information about these areas along with a number of other parameters are stored in the *Server of preferences*.

- **Location servers** receive encrypted location information from the m-Carer of the patients and store it. These servers could receive other data (*e.g.* heartbeat rate, body temperature, etc.). Note that all this information is encrypted and the server does not have access to the cryptographic keys to decrypt it. In this regard, location servers can be understood as log servers that store information of the whole system. The frequency at which m-Carers send information to the servers can be customised by the patients and the users of the system (*e.g.* Usual values range from one to ten minutes).

9.3 m-Carer operation and alarm states

From an operational perspective we can distinguish two main kinds of tasks: (i) administrative tasks, and (ii) monitoring tasks.

Administrative tasks are mainly related to the registration of users and patients into the system, the management of cryptographic keys, and the disclosure of information that requires the intervention of multiple parties. On the other hand, monitoring tasks are essentially focussed on the analysis of location information (and possibly some additional variables) with the aim to detect situations that could endanger the safety of the patients. In general, administrative tasks involve several actors (patients, users, servers, etc.) whilst monitoring tasks are essentially performed by the m-Carer running on the patients' devices.

In this section we describe two fundamental administrative tasks: the registration of patients and unofficial users (Section 9.3.1), and the disclosure of information (Section 9.3.5). Furthermore, we describe the three basic states of alarm (*i.e.* no alarm in Section 9.3.2, non-critical alarm in Section 9.3.3, and critical alarm in Section 9.3.4) and the operations performed to guarantee the privacy and safety of patients in each state.

9.3.1 Registration of patients and unofficial users

To illustrate the registration operation of our system, let us suppose that an old people's home is offering the service (*e.g.* in our running example this is The Golden Yard). A patient (*e.g.* Mr. Oldsmith) and an unofficial user (*e.g.* Ellie) agree with an official user (*e.g.* Mrs Andrews) that the patient is going to be under surveillance. Hence, the official user accesses the *Server of preferences* to register the new patient as well as one or more unofficial users related to the patient. During this procedure, the following actions take place:

1. The Server of preferences generates the cryptographic keys required by the system's protocols: a public/private key pair for the unofficial users $\{PK_{uu}, SK_{uu}\}$, a public/private key pair for the m-Carer running in the *Patient's Device* $\{PK_{pd}, SK_{pd}\}$ and a random encryption key K_{pd}. We assume that official users already have a public/private key pair $\{PK_{ou}, SK_{ou}\}$.

2. The initial conditions that lead to different alarm states are defined during this procedure (note that these conditions can be modified by the patient or the users at any time), as well as some other preferences. The following are the most significant parameters:

- *Secure areas.* As long as the patients remain in these secure/allowed areas no alarm will be raised regarding the location of the patient.
- *Non-allowed and dangerous areas.* When the patients enter those areas, the system raises an alarm. The type of alarm depends on the kind of area and it is fully customisable by the users.
- *Health parameters.* A number of gadgets could be attached to the Patient's Device (*e.g.* via Bluethooth) so as to measure, for instance, the heartbeat frequency or the body temperature. Users might assign different states of alarm for different values of these measures.
- *Frequency of messages.* This parameter defines how often a message containing location information (an possibly other) is sent to the *Location server*. Increasing this frequency benefits the resolution of the system at the cost of bandwidth usage.

3. Keys $\{SK_{pd}, PK_{uu}, PK_{ou}, K_{pd}\}$ and the m-Carer application are installed into the patient's device (*e.g.* the smartphone of Mr. Oldsmith).

4. Key PK_{pd} is sent to the *Location server* and it is associated to the *patient's device*.

5. Keys SK_{uu} and SK_{ou} are stored in the *Server of preferences*, protected by a password that is known only by the owners of the keys. Key K_{pd} is also stored in this server, and it is protected by a password known by all users related to the patient.

9.3.2 State of *No Alarm* (NA)

If the m-Carer running in the patient's device does not detect any of the alarm conditions defined, it periodically sends encrypted data to the *Location server*:

1. First, these data are encrypted with the public key of the *Unofficial user*, and the result is encrypted again this time with the public key of the *Official user*. In addition to these doubly encrypted data, the message contains a message authentication code (MAC):

$$\{\mathcal{ENC}_{PK_{ou}}(\mathcal{ENC}_{PK_{uu}}(data)), \mathcal{MAC}_{SK_{pd}}\}$$

2. The *Location server* stores the encrypted data and acknowledges the reception.

In our example, consider the locations 1 and 2 of the route of Mr. Oldsmith shown in Figure 9.4. In these points the m-Carer does not detect any alarm because Mr. Oldsmith is located in an allowed area that does not mean any risk to his safety. Hence, the m-Carer will take the current location of Mr. Oldsmith and will encrypt it first with the public key of Ellie and the result will be encrypted again with the public key of Mrs. Andrews. In this situation our system encrypts the location by using the RSA public key cryptosystem. However any other algorithm with the same security properties might be used.

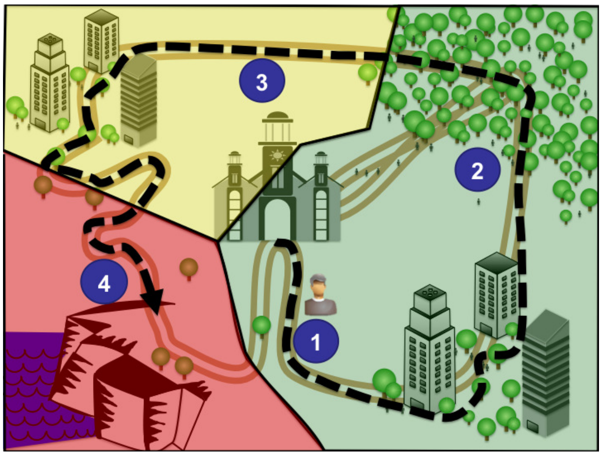

Figure 9.4: Route followed by Mr. Oldsmith in our example. Circled numbers indicate different moments in the promenade of Mr. Oldsmith that are used to illustrate different alarm states [106].

9.3.3 State of *Non-Critical Alarm* (NCA)

If the m-Carer detects a *Non-Critical* alarm condition, it sends an encrypted alert to the *Location server*:

1. The location along with other data are sent from the patient's device to the *Location server* using a *Non-Critical Alarm* message that contains the data encrypted with a symmetric key K_{pd}. A MAC is also added to the message to guarantee authenticity:

$$\{\mathcal{ENC}_{K_{pd}}(data), \mathcal{MAC}_{SK_{pd}}\}$$

2. Upon reception, the *Location server* forwards the message to all users related to the patient. Then, users can decrypt the message, since they know the symmetric key K_{pd}.

In our example, consider the location 3 in the route of Mr. Oldsmith shown in Figure 9.4. When the m-Carer analyses the location of Mr. Oldsmith at that point, it detects that he is located in a non-allowed area (marked in yellow). These areas do not represent a direct risk to the safety of the patient, but the fact that the patient is there might indicate that he is lost. In this situation the m-Carer sends the message described above to the *Location server* that forwards it to Ellie and Mrs. Andrews (*e.g.* they receive a warning message in their mobile phones). They can decrypt the message because they know the secret symmetric key K_{pd}. After decrypting the message Ellie and/or Mrs. Andrews may act accordingly. In this situation our system encrypts the location by using the 3DES symmetric key cryptosystem, however any other algorithm with, at least, the same security properties might be used (*e.g.* the Advanced Encryption Standard (AES) algorithm is an alternative).

9.3.4 State of *Critical Alarm* (CA)

If the m-Carer detects a critical situation (*e.g.* the patient is in a dangerous area, a fall has been detected, attached biosensors indicate a critical situation, etc.) it sends an unencrypted message to the *Location server* that forwards it to all users related to the patient. In addition, in parallel, the m-Carer sends a warning message to the *Emergency Services* informing about the location of the patient in clear text (*i.e.* no encryption is used).

In our example, consider the location number four of the route of Mr. Oldsmith shown in Figure 9.4. In this point, the m-Carer detects that Mr. Oldsmith is in a dangerous area (*i.e.* he is very close to a cliff). In this situation the safety of the patient is more important than his privacy, thus, the m-Carer automatically sends a warning message to the *Emergency services* (*e.g.* these can be public services or private services paid by The Golden Yard). Also, the m-Carer sends an unencrypted message with the location of Mr. Oldsmith to the *Location server* that immediately forwards it to Ellie and Mrs. Andrews. They will receive the warning in their mobile phones. In addition, the system can automatically make a distress voice call to warn them more reliably.

9.3.5 Obtaining information

As shown in the previous sections, depending on the alarm state, the m-Carer encrypts the location information differently. Thus, users have to use different procedures to decrypt the messages and obtain the information.

- If the message is sent under the *Critical Alarm* state, no encryption is used. Consequently, users receive the information in plain text and no further action is required on their side.

- In a *Non-Critical Alarm* state, messages are encrypted with a single symmetric key known by official and unofficial users related to the patient. Thus, they can individually decrypt the message by using the key K_{pd} as follows:

  $$\mathcal{DEC}_{K_{pd}}\{\mathcal{ENC}_{K_{pd}}(data)\} = data$$

  Note that only these users will be able to decrypt the message. The *Location server* that stores and forwards the messages is not able to decrypt them. In this regard, the privacy of the patient is fully guaranteed.

- For messages encrypted during normal operation (*i.e.* in a *No Alarm* state), official and unofficial users **have to collaborate** to decrypt the information. First, an official user partially decrypts the message using SK_{ou} and sends the result (partially decrypted) to an unofficial user that uses SK_{uu} to obtain the fully decrypted information.

$$\mathcal{DEC}_{SK_{ou}}\{\mathcal{ENC}_{PK_{ou}}(\mathcal{ENC}_{PK_{uu}}(data))\} = \mathcal{ENC}_{PK_{uu}}(data)$$

$$\mathcal{DEC}_{SK_{uu}}\{\mathcal{ENC}_{PK_{uu}}(data)\} = data$$

Note that neither official users nor unofficial users are able to obtain the information individually. Thus, the privacy of patients is guaranteed.

In our example, consider the location number two in the route of Mr. Oldsmith shown in Figure 9.4. Imagine that Ellie has to meet with this father but (for any reason) he has forgotten about the meeting. The m-Carer does not detect any risky situation because Mr. Oldsmith is located in an allowed safe area. Hence, it encrypts the location under the NA condition and sends it to the *Location server*. Ellie is worried and wants to know the location of her father. To do so, she contacts Mrs. Andrews and asks for her help to determine the location of Mr. Oldsmith. If they agree (let us assume that they do), Mrs. Andrews starts the decryption procedure by using her private key and sends the partially decrypted message to Ellie. Ellie receives this message and finalises the decryption using her private key and obtaining the location of his father.

Note that in this case in which no alarm is detected Mrs. Andrews (the official user) acts as a regulator to avoid the inappropriate use of the system. As a result the privacy of the Mr. Oldsmith is guaranteed.

9.4 Real implementation

In the previous sections we have described our m-Carer proposal from a theoretical perspective. In this section we provide a brief overview of a real implementation details of the system that is fully functional. This implementation has been done within the framework of "The LOST project" by a team of cryptographers and programmers of the CRISES Research Group at the Rovira i Virgili University in Spain. Note that this implementation is only a possibility. Several other options, based on the theoretical description given above, are also possible. Thus, the main goal of this section is not to deeply analyse the efficiency of the implementation but to show the feasibility of the proposal in the real world.

The communication protocols between users, patients and servers have been implemented in Java using a RESTful approach over HTTP. The web interface of the *Server of preferences* (shown in Figure 9.5) uses PHP, HTML and the Google Maps JavaScript API.

We have developed the m-Carer application that runs on the patient's device for the three main smartphones platforms, namely iOS, Android and Windows Mobile.

Regarding cryptography, we have used 3DES (Triple Data Encryption Standard) with 192-bit keys for the symmetric encryption and RSA with 1024-bit keys for the asymmetric/public-key encryption. Messages authentication codes (MAC) are obtained by using SHA-1 (Secure Hash Algorithm-1) and its RSA digital signature.

In addition, the battery life of the patients' devices has been tested in real life situations. For instance, using our m-Carer running on a Samsung Galaxy S with Android, the battery lasts for up to 62.5 hours (sending an encrypted location message every 5 minutes). Note that this result is only approximate and might vary

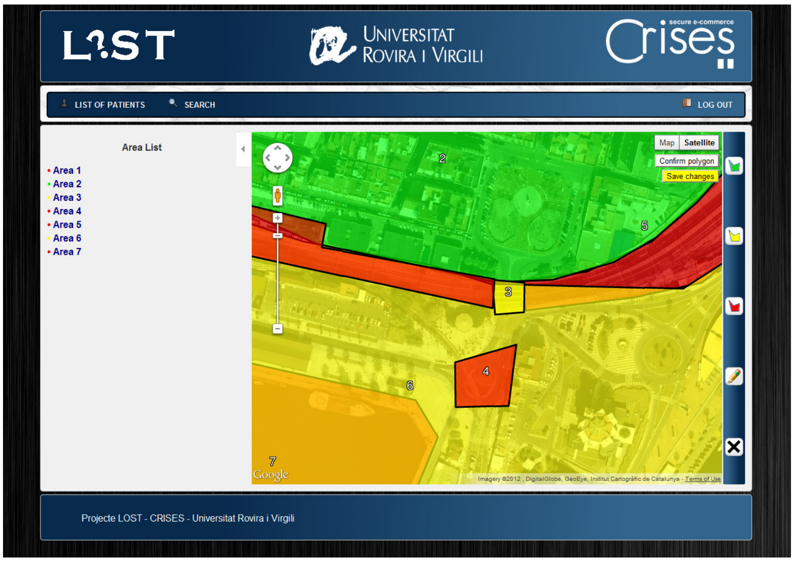
LOST
UNIVERSITAT ROVIRA I VIRGILI
secure e-commerce
Crises
LIST OF PATIENTS
SEARCH
LOG OUT
Area List
Area 1
Area 2
Area 3
Area 4
Area 5
Area 6
Area 7
Map
Satellite
Confirm polygon
Save changes
2
5
3
4
6
7
Google
Imagery ©2012 , DigitalGlobe, GeoEye, Institut Cartogràfic de Catalunya - Terms of Use
Projecte LOST - CRISES - Universitat Rovira i Virgili

Figure 9.5: Screenshot of the website used by users to define their preferences [106].

depending on the operating system, the GPS chipset, the data network, and the smartphone used.

Currently, we are planning and realising several trials in Tarragona and Barcelona (Catalonia) with the collaboration of the City Council, the University and local associations of relatives of patients with Alzheimer. The trials will last for, at least, six months and we expect to collect data that will allow us to improve the system and to better understand the wandering behaviour of patients with dementia.

Part IV

Conclusions and summary

Chapter 10
Summary of main contribution

In this chapter we present the findings and main contributions of this dissertation and the guidelines for future research in the studied topics.

Contents

10.1 Thesis Statement

Current technology has allowed the creation of new concepts such as Ubiquitous Computing and Intelligent Environments. Most people are connected remotely to a multitude of services that share data through mobile devices or sensors. The growth of these technologies has been so significant such that entire cities are based on a system where their services, infrastructures and citizens form a single fully connected environment.

These advances in technology have greatly facilitated the everyday lives of people, especially in sectors such as infrastructure, information, technology and health. But sometimes, especially in services where sensitive information is shared, through the advantages introduced by this technology, the risks of violating the privacy of its users is greater.

The aim of this dissertation is to study, analyse and propose new protocols in order to find new ways to improve the quality of life of citizens, through technology and efficient systems, whilst ensuring their privacy remains completely intact. In doing so, we have focused in three main areas:

- **Location-based services**
- **RFID and distributed sensors**

- **Smart Cities and Applications**

10.2 Location-based Services

Location-based services are gaining importance thanks to an increase in the number of mobile devices with self-location capabilities. Privacy plays a key role in the proper deployment of these services and it must be carefully considered. We have proposed a new distributed pseudonymizer to protect the privacy of the users of location-based services. Our proposal clearly improves the privacy level achieved by classic pseudonymizers (it does not rely on a TTP) whilst maintaining the same level of accuracy (the location is not distorted).

The complexity of the proposed method is low because only simple cryptographic operations are required. Also, the communications overhead is quite low and only depends on the number of users involved in the protocol.

Thanks to the avoidance of TTPs, our method: (i) scales properly; (ii) does not generate bottlenecks; (iii) does not require users to trust anybody; and (iv) does not have a single point of attack.

We have also introduced the concept of W^3-privacy related to the three dimensions of user privacy in LBS in chapter 3. We have also shown that most methods do not consider all dimensions and we have proposed a new method that combines the singular properties of prime numbers and privacy homomorphisms in order to preserve "query privacy" in chapter 4. Also, we indicate that combining our proposal with partially blind signatures and query k-anonymity, we can obtain a W^3-private location-based service.

10.3 RFID and distributed sensors

10.3.1 Radio Frequency Identification

Radio frequency identification (RFID) has become a must in the manufacturing and retailing sectors due to its ability to identify and track items rapidly and in parallel without the need for visual contact. This ability of the RFID technology has relegated bar codes to obsolescence. Moreover, RFID technology is especially suited for a variety of tasks, but due to the fact that RFID tags can be read by any reader in their range of cover, some security and privacy issues must be taken into account. Indeed, an eavesdropper could collect a great deal of information from citizens this way. Moreover, by making use of several readers strategically deployed in an area, it could be possible to track the location of people. So, if proper solutions are not taken into account, people could be profiled and tracked.

Usually, algorithms aimed at location prediction work well in some scenarios, and not as well in others. Although we have proposed the the use of the scalable method proposed in [112] in order to quick identify RFID tags embedded in cars within a smart city.

10.3.2 Distributed Sensors

Transportation in cities is particularly relevant for this topic of study. While, the management of public transportation is a complex issue for local governments. Private transportation has proved to be one of the most important focuses in big cities. The challenges related to private transportation are diverse, namely traffic jam management, tax collection, parking lot management, and so on.

With the aim of showing some of the privacy and security problems that might arise the use of sensors within a city, we have considered a case study focussed on managing parking payments. We have proposed a protocol that uses private enhancing techniques such as pseudonyms, improved randomised hash-locks and anonymous payments, to guarantee the privacy and security of the citizens that park their vehicles in public parking areas in a city. In chapter 7 we have discussed that our protocol allows anonymity, untraceability, remote payment and transparent multi-area parking. Further work includes the implementation of this protocol in a real scenario.

However, sensors in a city are not only used for infrastructure. The healthcare sector is struggling to provide better services whilst reducing costs. One of its main tools to achieve this goal is e-health, that allows the remote monitoring of patients, thereby reducing their hospitalisation time and costs. In addition, thanks to the wide use of mobile devices with computation capabilities, such as smart-phones and PDA, e-Health is transforming into m-Health, so that patients can be monitored on the move.

All these advances are possible thanks to the use of body sensors that can be attached to patients so as to collect data about their condition. Notwithstanding, body sensors can no longer be understood as isolated entities but, on the contrary, they are using ICT to create ad-hoc networks that allow them to communicate and exchange information.

The natural evolution of BSN is moving to store data in the cloud. In this way, it is possible to keep the computational and storage resources contained along with their cost. Also, thanks to the economies of scale related to the use of the cloud, BSN could grow easily and the sharing of information become simpler and much more efficient. However, the use of the cloud paradigm is not without risks since the data collected is no longer under personal control.

We have proposed a scheme based on the Raikova-Vo-Belkin-Malkin protocol and the new paradigm of Privacy as a Product in chapter 7. Our proposal guarantees the protection of the users' privacy whilst at the same time allowing the controlled sharing of information with multiple parties. We have shown that our protocol is feasible from a technical and economical perspective, and we have described several scenarios in which our proposal fits well and could be easily applied.

10.4 Smart Cities and Applications

10.4.1 Smart Cities

The concept of the Smart City has been adopted by many cities in the world and the challenge of becoming "smart" is gaining importance in the agenda of local governments. To be smart, cities must be sustainable, improve the quality of life of their citizens, foster their interaction through e-governance, etc. To achieve these goals, local governments are making serious efforts to move in the "smart direction" and private companies such as IBM and Cisco are playing (and will continue to play) a determinant role for the success of the smart cities of the future.

We believe that truly smart cities must protect the privacy of their citizens to be a true success. In chapter 8 we have presented the concept of Citizens Privacy. Our model distinguishes five dimensions, namely identity privacy, query privacy, location privacy, footprint privacy and owner privacy. Also, we have identified a number of real-life situations that might jeopardize the privacy of the citizens of a smart city and we have shown how to preserve privacy by using off-the-shelf Privacy Enhancing Technologies.

The technologies we have proposed are feasible and could be implemented in any smart city. However, their success will depend on some inherent aspects that should be addressed. For instance, the coexistence of multiple infrastructure domains should be properly tackled, and the transportation of the information between these parts should be achieved in a secure manner. Moreover, companies offering data center services for smart city infrastructure should take care of the security and reliability of their systems and networks. Finally, the adoption of security technologies in resource constrained devices –which is being solved thanks to the efforts of researchers and practitioners– must be also considered.

10.4.2 mHealth

The number of people with mild cognitive impairments grows and it will continue to grow in the future due to the ageing of our society. Most of these people are able to live a normal life, however, some can get lost or disoriented in some occasions. In these situations, they could be injured and finding them quickly is very important. Although some proposals to locate people exist, none of them fully consider the privacy of the user. This lack of privacy protection might prevent patients from choosing to use these systems. There is a clear need for new proposals that can balance the right of patients to be properly treated while maintaining their fundamental right to privacy.

Finally, we have proposed the new concept of the m-Carer in chapter 9. We have described a comprehensive architecture aimed at tracking people whilst guaranteeing their right to privacy. By using a set of personalised alarm states, relatives and human carers can easily find and help lost patients if necessary. With the aim of demonstrating the feasibility of our proposal, we have built a fully functional solution, and we have shown that our proposal is practical and useful. Although our proposal will help patients, human carers and relatives, we recognise that nowadays it is impossible to substitute entirely for a human carer with technology and there is still much work to be done in this area. The following are research lines that are going to be studied in the future so as to improve the m-Carer system proposed:

- Wandering detection: In this system patients can get lost within defined areas, without cutting off any alarms. It is desirable to develop methods to determine whether a patient is wandering even within safe areas however.

- Incremental learning of mobility patterns: The development of intelligent systems able to learn the movement patterns of patients might help to automatically adapt the boundaries of save and permitted areas in which patients can move.

- Integration with indoor systems: The concept of the m-Carer is basically designed for outdoors (due to the constraints of GPS receivers). Thus, integrating m-Carers with indoor monitoring systems would lead to a more comprehensive solution.

- Cellphone loss: It is possible that, patients might lose their cellphones. Determining the best way of carrying the cellphone to avoid its loss is still an open issue.

- Data integration: Cellphones are capable of gathering data from other devices via bluetooth or the like. However, the lack of well-known standards and the non-trivial nature of multiple sources data integration make this task very challenging and requires further consideration.

10.4.3 Publications

The main publications supporting the content of this thesis are the following:

- Pérez-martínez, Pablo A and Solanas, Agusti and Mart'ınez-Ballesté, Antoni. Location Privacy Through Users' Collaboration: A Distributed Pseudonymizer. In *Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, 2009. UBICOMM'09.* Pages 338–341, 2009.

- Pérez-Martínez, Pablo A and Solanas, Agusti. W3-Privacy: the Three Dimensions of User Privacy in LBS. In *MobiHoc 2011, the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing*

- Martinez-Balleste, Antoni and Pérez-martínez, Pablo A and Solanas, Agusti. The pursuit of citizens' privacy: a privacy-aware smart city is possible. In *Communications Magazine, IEEE.* Volume 51, pages 136–141, 2013.

- Solanas, Agusti and Mart'ınez-Ballesté, Antoni and Pérez-Martínez, Pablo A and Pena, Albert Fernandez de la and Ramos, Javier. m-carer: Privacy-aware monitoring for people with mild cognitive impairment and dementia. In *IEEE Journal on Selected Areas in Communications.* Volume 31, number 9, pages 19–27, 2013.

- Pérez-Martínez, Pablo A and Martínez-Ballesté, Antoni and Solanas, Agusti. Privacy in Smart Cities-A Case Study of Smart Public Parking. In *PECCS - International Conference on Pervasive and Embedded Computing and Communication Systems.* Pages 55–59, 2013.

Other publications co-authored by the candidate and related to privacy protection, but not included in this thesis, are listed below:

- Solanas, Agusti and Patsakis, Constantinos and Conti, Marco and Vlachos, Ioannis and Ramos, Victoria and Falcone, Francisco and Postolache, Octavian and Pérez-martínez, Pablo A and Pietro, Roberto and Perrea, Despina and Martínez-Ballesté, Antoni. Smart health: a context-aware health paradigm within smart cities. In *Communications Magazine, IEEE.* Volume 52, number 8, pages 74–81, 2014.

- Trujillo-Rasua, Rolando and Solanas, Agusti and Pérez-Martínez, Pablo A and Domingo-Ferrer, Josep. Predictive protocol for the scalable identification of RFID tags through collaborative readers. In *Computers in Industry.* Volume 63, number 6, pages 557–573, 2012.

Bibliography

[1] Aol. http://www.aol.com.

[2] Boycott benetton. http://www.boycottbenetton.com.

[3] Verichip corporation, 2005. http://www.4verichip.com.

[4] Masayuki Abe and Tatsuaki Okamoto. Provably secure partially blind signatures. In *Proceedings of the 20th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '00, pages 271–286, London, UK, 2000. Springer-Verlag.

[5] American Civil Liberties Union (ACLU). Rfid position statement of consumer privacy and civil liberties organitzations. http://www.privacyrights.org/ar/RFIDposition.htm.

[6] Barcelona Activa. Live barcelona. Website, 2012. http://w41.bcn.cat/web/guest.

[7] Hande Alemdar and Cem Ersoy. Wireless sensor networks for healthcare: A survey. *Computer Networks*, 54(15):2688 – 2710, 2010.

[8] Jordi Aragones-Vilella, Antoni Martinez-Balleste, and Agusti Solanas. A brief survey on RFID privacy and security. *Proceedings of the World Congress on Engineering 2007*, 2:1488–1493, 2007.

[9] C. A. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and P. Samarati. Location privacy protection through obfuscation-based techniques. In S. Baker and G.J. Ahn, editors, *Data and Applications Security*, volume 4602 of *LNCS*, pages 47 – 60. IFIP, 2007.

[10] Gildas Avoine, Levente Buttyant, Tamas Holczer, and Istvan Vajda. Group-based private authentication. *In IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2007)*, pages 1–6, 2007.

[11] Gildas Avoine, Etienne Dysli, and Philippe Oechslin. Reducing time complexity in RFID systems. *In Bart Preneel and Stafford Tavares, editors,* Selected Areas in Cryptography*, volume 3897 of Lecture Notes in Computer Science*, pages 291–306, 2006.

[12] Feng Bao and Robert Deng. Privacy protection for transactions of digital goods. In Sihan Qing, Tatsuaki Okamoto, and Jianying Zhou, editors, *Information and Communications Security*, volume 2229 of *Lecture Notes in Computer Science*, pages 202–213. Springer Berlin Heidelberg, 2001.

[13] Victoria Barret. Dropbox: The inside story of tech's hottest startup. http://www.forbes.com/sites/victoriabarret/2011/10/18/dropbox-the-inside-story-of-techs-hottest-startup/, Nov. 7, 2011.

[14] Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2:46–55, January 2003.

[15] Claudio Bettini, X. Wang, and Sushil Jajodia. Protecting privacy against location-based personal identification. In Willem Jonker and Milan Petkovic, editors, *Secure Data Management*, volume 3674 of *Lecture Notes in Computer Science*, pages 185–199. Springer Berlin / Heidelberg, 2005.

[16] Burton H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM*, 13(7):422–426, July 1970.

[17] Maged N Kamel Boulos, Artur Rocha, Angelo Martins, Manuel Escriche Vicente, Armin Bolz, Robert Feld, Igor Tchoudovski, Martin Braecklein, John Nelson, Gearoid O Laighin, Claudio Sdogati, Francesca Cesaroni, Marco Antomarini, Angela Jobes, and Mark Kinirons. Caalyx: a new generation of location-based services in healthcare. *International Journal of Health Geographics*, 6(9):1–6, mar 2007.

[18] Mark Burgess. An approach to understanding policy based on autonomy and voluntary cooperation. In *16th IFIP/IEEE Distributed Systems Operations and Management (DSOM 2005), LNCS 3775*, pages 97–108. Springer-Verlag, 2005.

[19] Mark Burgess and G. Canright. Scalability of peer configuration management in logically ad hoc networks. *Network and Service Management, IEEE Transactions on*, 1(1):21–29, 2004.

[20] Expatica Communications BV. Waiting for the bus that will never come, June 2008. http://www.expatica.com/de/health_fitness/healthcare/Waiting-for-the-bus-that-will-never-come.html.

[21] Christian Cachin, Silvio Micali, and Markus Stadler. Computationally private information retrieval with polylogarithmic communication. In *Advances in Cryptology EUROCRYPT '99*, pages 402–414. Springer-Verlag, 1999.

[22] Andrea Caragliu, Chiara del Bo, and Peter Nijkamp. Smart cities in europe. In *CERS'09, 3rd Central European Conference in Regional Science*, pages 45 – 59, October 2009.

[23] Roberto Casas, Alvaro Marco, Jorge L. Falco, Hector Gracia, and Alvaro Marco. Dalma - location aware alarm system for people with disabilities. In *Computers Helping People with Special Needs*, volume 4061 of *Lecture Notes in Computer Science*, pages 744–751. Springer, Berlin - Heidelberg, 2006.

[24] Dan Chalmers, Matthew Chalmers, Jon Crowcroft, Marta Kwiatkowska, Robin Milner, Eammon O'Neill, Tom Rodden, Vladimiro Sassone, and Morris Sloman. Ubiquitous computing: Experience, design and science. Technical report, 2006.

[25] Marinos Charalambides, Paris Flegkas, George Pavlou, Arosha K B, and Emil C Lupu. Policy conflict analysis for quality of service management. In *In Proceedings 6th IEEE Workshop on Policies for Distributed Systems and Networks (Policy 2005*, pages 99–108. IEEE Computer Society, 2005.

[26] Marinos Charalambides, Paris Flegkas, George Pavlou, and Javier Rubio-loyola. Dynamic policy analysis and conflict resolution for diffserv quality of service management. In *in Proceedings of the IEEE/IFIP Network Operations and Management Symposium 2006*, pages 294–304, 2006.

[27] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan. Private information retrieval. In *Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on*, pages 41 –50, oct 1995.

[28] Benny Chor and Niv Gilboa. Computationally private information retrieval (extended abstract). In *Journal of the ACM*, pages 41–50, 1997.

[29] ChiYin Chow, Mohamed F. Mokbel, and Xuan Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based services. In *GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 171–178. ACM, November 2006.

[30] Boyd Cohen. The top 10 smart cities on the planet. On-line at http://www.fastcompany.com, 2012. http://www.fastcoexist.com/1679127/the-top-10-smart-cities-on-the-planet.

[31] European Commission. Protection of personal data. On-line at http://ec.europa.eu/justice/data-protection/index_en.htm, 2013.

[32] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *In Proceedings of the 13th USENIX Security Symposium*, pages 303–320, 2004.

[33] Josep Domingo, Agusti Solanas, and Jordi Castellà. h(k)-private information retrieval from privacy-uncooperative queryable databases. *Online Information Review*, 33(4):720–744, 2009.

[34] J. Domingo-Ferrer, M. Sramka, and R. Trujillo-Rasúa. Privacy-preserving publication of trajectories using microaggregation. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*, pages 26–33. ACM, 2010.

[35] Josep Domingo-Ferrer. A three-dimensional conceptual framework for database privacy. In Willem Jonker and Milan Petkovic, editors, *Secure Data Management*, volume 4721 of *Lecture Notes in Computer Science*, pages 193–202. Springer, 2007.

[36] Matt Duckham and Lars Kulik. A formal model of obfuscation and negotiation for location privacy. In Hans W. Gellersen, Roy Want, and Albrecht Schmidt, editors, *Pervasive Computing*, volume 3468 of *Lecture Notes in Computer Science*, pages 243–251. Springer Berlin / Heidelberg, 2005.

[37] Cynthia Dwork. Differential privacy. In *ICALP: Annual International Colloquium on Automata, Languages and Programming*, 2006.

[38] T. Elgamal. A public key cryptosystem and a signature scheme based on discrete logarithms. *Information Theory, IEEE Transactions on*, 31(4):469 – 472, jul 1985.

[39] K. Fishin, S. Roy, B. Jiang, Kenneth P. Fishkin, Sumit Roy, and Bing Jiang. Some methods for privacy in rfid communication, 2004.

[40] Sepideh Fouladgar and Hossam Afifi. Scalable privacy protecting scheme through distributed RFID tag identification. *In Proceedings of the workshop on Applications of private and anonymous communications*, 3:1–8, 2008.

[41] B. Gedik and L. Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, 7(1):1–18, 2008.

[42] Bugra Gedik and Ling Liu. Location privacy in mobile systems: A personalized anonymization model. In *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems*, ICDCS '05, pages 620–629, Washington, DC, USA, 2005. IEEE Computer Society.

[43] Buğra Gedik and Ling Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, 7:1–18, January 2008.

[44] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 169–178, New York, NY, USA, 2009. ACM.

[45] P. Germanakos, C. Mourlas, and G. Samaras. A mobile agent approach for ubiquitous and personalized ehealth information systems. In *Proc. Workshop on 'Personalization for e-Health' of the 10th International Conference on User Modeling (UM'05). Edinburgh*, pages 67–70, 2005.

[46] Yael Gertner, Yuval Ishai, Eyal Kushilevitz, and Tal Malkin. Protecting data privacy in private information retrieval schemes. In *Proceedings of the thirtieth*

annual ACM symposium on Theory of computing, STOC '98, pages 151–160, New York, NY, USA, 1998. ACM.

[47] G. Ghinita, P. Kalnis, S. Skiadopoulus, and J. Joxan. Prive: Anonymous location-based queries in distributed mobile systems. In *In WWW '07: Proceedings of the 16th International Conference on World Wide Web*, pages 371–380. ACM Press, 2007.

[48] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan. Private queries in location based services: anonymizers are not necessary. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, SIGMOD '08, pages 121–132, New York, NY, USA, 2008. ACM.

[49] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan. Private queries in location based services: Anonymizers are not necessary. In *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 121 – 132. Vancouver, BC, Canada, ACM, June 2008.

[50] Gabriel Ghinita, Panos Kalnis, and Spiros Skiadopoulos. Mobihide: A mobile peer-to-peer system for anonymous location-based queries, 2007.

[51] Inc. Google. Google, inc. http://www.google.com.

[52] Inc. Google. Google maps. http://maps.google.com.

[53] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proc. MobiSys'03*, pages 31 – 42, 2003.

[54] Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services*, MobiSys '03, pages 31–42, New York, NY, USA, 2003. ACM.

[55] A.M. Hadjiantonis, M. Charalambides, and G. Pavlou. A policy-based approach for managing ubiquitous networks in urban spaces. In *Communications, 2007. ICC '07. IEEE International Conference on*, pages 2089–2096, 2007.

[56] Y. Hao and R. Foster. Wireless body sensor networks for health-monitoring applications. *Physiological measurement*, 29(11):R27, 2008.

[57] R.R. Heckle. Security dilemma: Healthcare clinicians at work. *Security & Privacy, IEEE*, 9(6):14 –19, nov.-dec. 2011.

[58] Eric Hughes. A cypherpunk's manifesto. http://www.activism.net/cypherpunk/manifesto.html, 1993.

[59] Julian C Hughes and Stephen J Louw. Electronic tagging of people with dementia who wander. *British Medical Journal*, 325:847–848, 2002.

[60] Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Eric Schulte Nordholt, Keith Spicer, and Peter-Paul de Wolf. *Statistical Disclosure Control*. John Wiley & Sons, Ltd, 2012.

[61] IBM. Integrated fare management for transportation. Website, 2011. http://www.ibm.com/smarterplanet/us/en/traffic_congestion/nextsteps/solution/G080151085496M88.html.

[62] Ari Juels. Rfid security and privacy: A research survey. *Journal of Selected Areas in Communication*, 24(2):381–395, 2006.

[63] Ari Juels, Ronald L. Rivest, and Michael Szydlo. The blocker tag: Selective blocking of rfid tags for consumer privacy. In *8th ACM Conference on Computer and Communications Security*, pages 103–111. ACM Press, 2003.

[64] Ari Juels and Stephen A. Weis. Defining strong privacy for RFID., 2006.

[65] Waleed K. A. Abdulrahem, Hebah H. O. Nasereddin, and Said M. H. Fares. Business continuity based on rfid. volume 5. American Academic and Scholarly Research Center, 4 2013.

[66] Eija Kaasinen. User needs for location-aware mobile services. *Personal and Ubiquitous Computing*, 7(1):70–79, 2003.

[67] Ali Khoshgozaran, Houtan Shirani-Mehr, and Cyrus Shahabi. Spiral: A scalable private information retrieval approach to location privacy. *Mobile Data Management Workshops, 2008 Ninth International Conference on*, 0:55–62, 2008.

[68] Eyal Kushilevitz and Rafail Ostrovsky. Replication is not needed: Single database, computationally-private information retrieval (extended abstract). In *In proceedings of the 38th annual IEEE Symposioum on Foundations of Computer Science*, pages 364–373, 1997.

[69] S. Lee, D. Yoon, and A. Ghosh. Intelligent parking lot application using wireless sensor networks. In *Collaborative Technologies and Systems, 2008. CTS 2008. International Symposium on*, may 2008.

[70] Liander and AIM. Amsterdam smart city. Website, 2012. http://www.amsterdamsmartcity.nl/#/en.

[71] E. Lupu and M. Sloman. Conflicts in policy-based distributed systems management. *IEEE Transactions on Software Engineering*, 25:852–869, 1999.

[72] Alvaro Marco, Roberto Casas, Jorge L. Falco, Hector Gracia, Jose I. Artigas, and A. Roy. Location-based services for elderly and disabled people. *Computer Communications*, 31(6):1055–1066, 2008.

[73] A. Martinez-Balleste, P.A. Perez-Martinez, and A. Solanas. The pursuit of citizens' privacy: a privacy-aware smart city is possible. *Communications Magazine, IEEE*, 51(6):136–141, June 2013.

[74] Antoni Martínez-Ballesté, Hatem A. Rashwan, Domenec Puig, and Antonia Paniza Fullana. Towards a trustworthy privacy in pervasive video surveillance systems. In *PerCom Workshops*, pages 914–919. IEEE, 2012.

[75] R McShane, K Gedling, J Keene, C Fairburn, R Jacoby, and T Hope. Getting lost in dementia: a longitudinal study of a behavioral symptom. *International Psychogeriatrics*, 10(3):253–260, 1998.

[76] Medical-Intelligence. Columba bracelet. http://www.medicalintelligence.ca/en/products/hardwares/prima/tech.html.

[77] M. Meingast, T. Roosta, and S. Sastry. Security and privacy issues with health care information technology. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 5453 –5458, 30 2006-sept. 3 2006.

[78] Frank Miskelly. Electronic tracking of patients with dementia and wandering using mobile phone technology. *Age and Ageing*, 34:497–518, 2005.

[79] Jonathan D. Moffett and Morris S. Sloman. Policy conflict analysis in distributed system management, 1993.

[80] N. Mohammed, B. Fung, and M. Debbabi. Walking in the crowd: anonymizing trajectory data for pattern analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 1441–1444. ACM, 2009.

[81] Mohamed F. Mokbel, Chi-Yin Chow, and Walid G. Aref. The new casper: query processing for location services without compromising privacy. In *Proceedings of the 32nd international conference on Very large data bases*, VLDB '06, pages 763–774. VLDB Endowment, 2006.

[82] David Molnar, Andrea Soppera, and David Wagner. Privacy for rfid through trusted computing. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, WPES '05, pages 31–34, New York, NY, USA, 2005. ACM.

[83] David Molnar and David Wagner. Privacy and security in library RFID: issues, practices, and architectures. *In Proceedings of the 11th ACM conference on Computer and communications security*, pages 210–219, 2004.

[84] HS Ng, ML Sim, and CM Tan. Security issues of wireless sensor networks in healthcare applications. *BT Technology Journal*, 24(2):138–144, 2006.

[85] G. Ostojic, S. Stankovski, M. Lazarevic, and V. Jovanovic. Implementation of RFID technology in parking lot access control system. In *RFID Eurasia, 2007 1st Annual*, pages 1 –5, sept. 2007.

[86] C. Otto, A. Milenkovic, C. Sanders, and E. Jovanov. System architecture of a wireless body area sensor network for ubiquitous health monitoring. *Journal of Mobile Multimedia*, 1(4):307–326, 2006.

[87] Vasilis Pappas, Mariana Raykova, Binh Vo, Steven M. Bellovin, and Tal Malkin. Private search in the real world. In Robert H'obbes' Zakon, John P. McDermott, and Michael E. Locasto, editors, *ACSAC*, pages 83–92. ACM, 2011.

[88] Constantinos Patsakis and Agusti Solanas. Privacy as a product: A case study in the m-health sector. *Special issue of Journal of Network and Computer Applications, in Collaborative Technologies and Applications*, 2012.

[89] Sai Teja Peddinti and Nitesh Saxena. Web search query privacy: Evaluating query obfuscation and anonymizing networks. *J. Comput. Secur.*, 22(1):155–199, January 2014.

[90] People-Track-USA. Gps monitoring services for alzheimer patients and children with autism. http://www.peopletrackusa.com/GPSTrackingofAlzheimerPatientsandAutismWanderers.html.

[91] P.A. Pérez-Martínez and A. Solanas. W3-privacy: the three dimensions of user privacy in lbs. In *MobiHoc 2011, the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2011.

[92] Pablo A. Pérez-Martínez, A. Martínez-Ballesté, and A. Solanas. Privacy in smart cities: A case study of smart public parking. *Third International Conference on Pervasive and Embeded Computing and Communications Systems, PECCS 2013*, 2012.

[93] Pablo A. Perez-Martinez, Agusti Solanas, and Antoni Martinez-Balleste. Location privacy through users' collaboration: A distributed pseudonymizer. *Mobile Ubiquitous Computing, Systems, Services and Technologies, International Conference on*, 0:338–341, 2009.

[94] Pablo A. Pérez-Martínez, Agusti Solanas, and Antoni Martínez-Ballesté. Location privacy through users' collaboration: A distributed pseudonymizer. In *Proceedings of the 2009 Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, UBICOMM '09, pages 338–341, 2009.

[95] RC Petersen, R Doody, A Kurz, RC Mohs, JC Morris, PV Rabins, K Ritchie, M Rossor, L Thal, and B Winblad. Current concepts in mild cognitive impairment. *Archives of Neurology*, 58(12):1985–1992, 2001.

[96] P. Ray, N. Parameswaran, V. Chan, and W. Yu. Awareness modelling in collaborative mobile e-health. *Journal of Telemedicine and Telecare*, 14(7):381–385, 2008.

[97] Mariana Raykova, Binh Vo, Steven M. Bellovin, and Tal Malkin. Secure anonymous database search. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, CCSW '09, pages 115–126, New York, NY, USA, 2009. ACM.

[98] MelanieR. Rieback, Bruno Crispo, and AndrewS. Tanenbaum. Keep on blockin' in the free world: Personal access control for low-cost rfid tags. In Bruce Christianson, Bruno Crispo, JamesA. Malcolm, and Michael Roe, editors, *Security Protocols*, volume 4631 of *Lecture Notes in Computer Science*, pages 51–59. Springer Berlin Heidelberg, 2007.

[99] Tom Rodden, Adrian Friday, Henk Muller, and Alan Dix. A lightweight approach to managing privacy in location-based services, equator-02-058. Technical Report CSTR-07-006, University of Nottingham and Lancaster University and University of Bristol, October 2002.

[100] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.

[101] H. Schulzrinne, H. Tschofenig, J. Morris, J. Cuellar, and J. Polk. Geolocation policy. Technical report, Internet Engineering Task Force, June 2008. `http://www.ietf.org/internet-drafts/draft-ietf-geopriv-policy-17.txt`.

[102] Loren Schwiebert, Sandeep K.S. Gupta, and Jennifer Weinmann. Research challenges in wireless networks of biomedical sensors. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, MobiCom '01, pages 151–165, New York, NY, USA, 2001. ACM.

[103] Daniel R. Simon. Untraceable electronic cash. Patent US005768385A, August 1995.

[104] Einar Snekkenes. Concepts for personal location privacy policies. In *ACM Conference on Electronic Commerce*, pages 48–57, 2001.

[105] A. Solanas and A. Martínez-Ballesté. A TTP-free protocol for location privacy in location-based services. *Computer Communications*, 31(6):1181–1191, 2008.

[106] A. Solanas, A. Martínez-Ballesté, Pablo A. Pérez-Martínez, A. Fernandez, and J. Ramos. m-carer: Privacy-aware monitoring for people with mild cognitive impairment and dementia. *IEEE Journal on Selected Areas in Communications*, 2013.

[107] A. Solanas, C. Patsakis, M. Conti, I. Vlachos, V. Ramos, F. Falcone, O. Postolache, P. Perez-martinez, R. Pietro, D. Perrea, and A. Martinez-Balleste.

Smart health: A context-aware health paradigm within smart cities. *Communications Magazine, IEEE*, 52(8):74–81, Aug 2014.

[108] Agusti Solanas, Josep Domingo-Ferrer, Antoni Martínez-Ballesté, and Vanesa Daza. A distributed architecture for scalable private RFID tag identification. *Computer Networks*, 51(9):2268–2279, 2007.

[109] Agusti Solanas and Antoni Martínez-Ballesté. Privacy protection in location-based services through a public-key privacy homomorphism. In *EuroPKI*, pages 362–368, 2007.

[110] Vodafone (Spain). Simap. http://www.simapglobal.com/.

[111] M. Terrovitis and N. Mamoulis. Privacy preservation in the publication of trajectories. In *Mobile Data Management, 2008. MDM'08. 9th International Conference on*, pages 65–72. IEEE, 2008.

[112] R. Trujillo-Rasua, A. Solanas, Pablo A. Pérez-Martínez, and J. Domingo-Ferrer. Predictive protocol for the scalable identification of rfid tags through collaborative readers. *Computers in Industry*, 63(6):557–573, 2012.

[113] Rolando Trujillo-Rasua and Agusti Solanas. Efficient probabilistic communication protocol for the private identification of RFID tags by means of collaborative readers. *Computer Networks*, 55(15):3211–3223, 2011.

[114] Rolando Trujillo-Rasua and Agusti Solanas. Scalable trajectory-based protocol for RFID tags identification., 2011.

[115] J. Vaidya and C. Clifton. Privacy-preserving data mining: why, how, and when. *Security Privacy, IEEE*, 2(6):19 – 27, nov.-dec. 2004.

[116] Vital Wave Consulting. mhealth for development: The opportunity of mobile technology for healthcare in the developing world. Technical report, UN Foundation-Vodafone Foundation Partnership, 2009.

[117] Qianhong Wu, Jordi Castellà, Josep Domingo, Agusti Solanas, and Bo Qin. Portable device-centric mutual privacy in priced location-based services. Manuscript.

[118] G.Z. Yang and M. Yacoub. *Body sensor networks*, volume 6. Springer Berlin, Germany:, 2006.

[119] Lin Yao, Chi Lin, Guangya Liu, Fangyu Deng, and Guowei Wu. Location anonymity based on fake queries in continuous location-based services. In *Availability, Reliability and Security (ARES), 2012 Seventh International Conference on*, pages 375–382, Aug 2012.

[120] X. Yi, M. Kaosar, R. Paulet, and E. Bertino. Single-database private information retrieval from fully homomorphic encryption. *Knowledge and Data Engineering, IEEE Transactions on*, PP(99):1, 2012.

[121] Man Lung Yiu, Christian S. Jensen, Xuegang Huang, and Hua Lu. Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. In *IEEE 24th International Conference on Data Engineering ICDE'08*, pages 366–375, 2008.