The Pursuit of Citizens Privacy: A Privacy-Aware Smart City is Possible

Antoni Martínez-Ballesté, Pablo A. Pérez-Martínez, Agusti Solanas, *Member, IEEE*

Abstract—Cities are growing steadily and the process of urbanization is a common trend in the world. Although cities are getting bigger, they are not necessarily getting better. With the aim to provide citizens with a better place to live, a new concept of city was born: the Smart City.

The real meaning of smart city is not strictly defined but it has gained much attention and many cities are taking action so as to be considered "smart". These smart cities, founded on the use of Information and Communication Technologies, aim at tackling many local problems, from local economy and transportation to quality of life and e-governance. Although technology helps to solve many of these local problems, their ability of gathering unprecedented amounts of information could endanger the privacy of the citizens.

In this article we identify a number of privacy breaches that can appear within the context of smart cities and their services. We leverage some concepts of previously defined privacy models and 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.

I. 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 [9] 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 [8], 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 [2].

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 [3] within a smart city would be the icing on the cake.

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

In this article, 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.
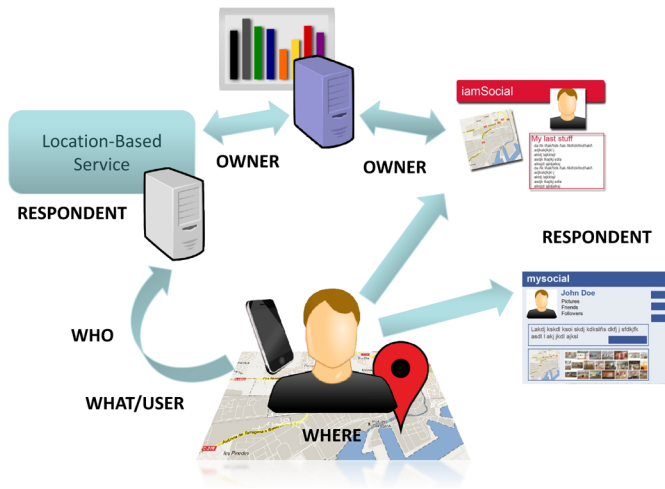
Fig. 1. Conceptual scheme of the privacy models. In the picture we distinguish (i) a user (at the bottom) that contacts a location-based service provider, (ii) two social networks (on the right) to which the user belongs, and (iii) a data warehousing facility (at the top).

II. BACKGROUND ON 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 [4] and the W^3-(Where, Who, What) privacy model for location-based services (LBS) [11]. Figure 1 illustrates the theoretical privacy dimensions described in this section.

A. 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 [4] 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 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 III below.

B. W^3-Privacy for Location-Based Services

Services related to the location of the user are gaining importance and so do privacy issues related to them. In [11] Pérez and Solanas 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.

The authors of [11] affirm that 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.

III. 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 IV.

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 [7]. 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* [5].

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 [1]. 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 [16]. However, in general PIR approaches are considered to be impractical in real scenarios yet.

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 [15]. 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.

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 [6] 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 [14].

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 [12].

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 [13]).

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 [10], 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.

IV. 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 III) that could be used to address those concerns. The examples used throughout the section refer to Figure 2. The goal of the scenario depicted in Figure 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.
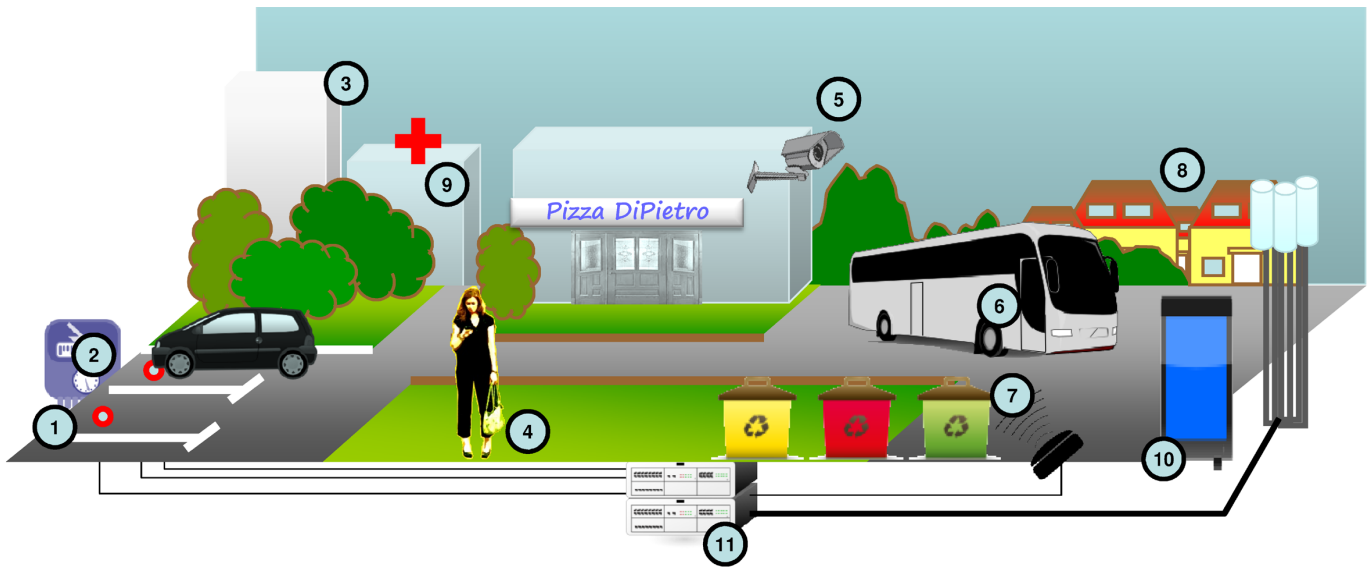
Fig. 2. Our example of smart city: (1) Smart parking service (2) Electric car recharging (3) Smart building with control of presence (4) User querying an LBS provider (5) Camera for video surveillance (6) Smart bus that changes the route upon users needs (7) Smart garbage containers (8) Control of energy consumption in homes (9) Medical center (10) Interactive information pole (11) Network infrastructure of the smart city.

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 I.

A. 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.

B. 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.

TABLE I
SUMMARY OF OUR 5D PROPOSAL FOR MODELING THE PRIVACY ASPECTS IN SMART CITY SERVICES.

Our 5D approach	Mapping to existing models		Examples of privacy concerns	Existing solutions
	3D Database	W3 LBS		
Identity		Who	Most of the examples entail identity privacy concerns. RFID and videosurveillance 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

Using TTP might also be an option. Whatever technique is applied, the goal should be hampering the correlation of users and queries.

C. 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.

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.

D. 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.

E. 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.

V. CONCLUSIONS

The concept of Smart City has been adopted by many cities in the world and the challenge of being "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, an so on. To achieve these goals, local governments are making serious efforts to move in the "smart direction" and private companies like IBM and Cisco are playing (and will play) a determinant role for the success of the smart cities of the future.

We believe that real smart cities count with their citizens and they must protect their privacy to make this challenge a true success. In this article 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 it 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 done in a secure manner. Moreover, companies offering data center services for the 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.

REFERENCES

[1] 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.

[2] 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.

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

[4] 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.

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

[6] 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.

[7] 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.

[8] IBM. Integrated fare management for transportation. Website, 2011. http://www.ibm.com/smarterplanet/us/en/traffic_congestion/nextsteps/solution/G080151O85496M88.html.

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

[10] 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.

[11] Pablo A. Pérez-Martínez and Agusti Solanas. W3-privacy: the three dimensions of user privacy in LBS. In *Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing - MOBIHOC 2011, Paris, France, May 2011*. ACM, 2011.

[12] 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.

[13] 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.

[14] Agusti Solanas and Antoni Martínez-Ballesté. A TTP-free protocol for location privacy in location-based services. *Computer Communications*, 31(6):1181–1191, 2008.

[15] J. Vaidya and C. Clifton. Privacy-preserving data mining: why, how, and when. *Security Privacy, IEEE*, 2(6):19 – 27, nov.-dec. 2004.

[16] 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.