AMORES L1.2 - A Privacy Risk Assessment Methodology for Location-Based Systems

Jesús Friginal, Jérémie Guiochet, Marc-Olivier Killijian

Livrable AMORES

AMORES L1.2 - A Privacy Risk Assessment Methodology for Location-Based Systems

Date	:	26 mars 2014
Auteurs	:	Jesús Friginal, Jérémie Guiochet, Marc-Olivier Killijian
Titre	:	AMORES L1.2 - A Privacy Risk Assessment Methodology for Location-Based Systems
Rapport No. / Version	:	1.2/ 1.0
Statut	:	Final version

LAAS-CNRS
7 avenue du Colonel Roche
31077 Toulouse cedex 4, France
http://www.laas.fr/

Supélec - Rennes
Avenue de la Boulaie, BP 81127,
35511 Cesson-Sévigneé, France
http://www.supelec.fr/

IRISA
CNRS, INRIA, Université de Rennes 1,
263, Avenue du Général Leclerc, Bat12
35042 Rennes Cedex, France
http://www.irisa.fr/

MobiGIS
Rue de l'Autan
31330 Grenade, France
http://www.mobigis.fr/

Tisséo
7, esplanade Compans-Caffarelli
31902 Toulouse CEDEX 9
http://www.tisseo.fr/

Executive summary

The proposed process is based on classical risk assessment process as defined in the risk management process. We adapt the different steps to the domain of geo-privacy, and we propose a meta-model of concepts used for this study. This meta-model has been successfully used to compute privacy risk and its relation with attack trees. The feasibility of each step of our methodology is illustrated through the case study of dynamic carpooling of the AMORES project (see report L.1.1.a).

1 Introduction

After the successful development of positioning technology, such as GPS, and the rise of infrastructureless wireless networks, such as ad hoc networks, mobile and ubiquitous (mobiquitous) systems have become the spearhead sector of the communication industry. The vast deployment of myriads of sensors and the rapid growth in the number of mobile devices per person is providing enormous opportunities to create a new generation of innovative Location-Based Services (LBS) addressed to improve the welfare of our society. LBS are used in a variety of contexts, such as health, entertainment or work, like discovering the nearest cash machine, parking parcel, or getting personalised weather services.

Parallel to this revolution, numerous studies reveal potential privacy breaches in the use of these services given the sensitivity of the collected information, and how it is stored and exchanged [GKNndPC11]. From a privacy viewpoint, the main characteristic of LBS systems is that, apart from personal identity, users' location becomes a new essential asset to protect. A recent study from MIT [dMHVB13] showed that 4 spatio-temporal points (approximate places and times), were enough to unequivocally identify 95% people in a mobility database of 1.5M users. The study shows that these constraints hold even when the resolution of the dataset is low. Therefore, even coarse or blurred information provides little anonymity. However, very few users are aware of the implications that a misuse of their location information may have on their privacy, and the potential consequences on their security and safety [DF03]. Tackling this issue becomes critical given the increasingly variety of attacks that may impact the privacy of users [GKNndPC11].

By the time being, there is a range from simplistic on/off switches to sophisticated Privacy-Enhancing Technologies (PETs) using anonymisation techniques [MCA06]. Today, few LBS offer such PETs, e.g., Google Latitude offers an on/off switch that allows to stick one's position to a freely definable location. Another set of techniques include location obfuscation, which slightly alter the location of the users in order to hide their real location while still being able to represent their position and receive services from their LBS provider. However, such efforts remain questionable in practice while suitable techniques to guarantee acceptable levels of risk remain unavailable. Consequently, the confident use of LBS requires not only the development of PETs, but also the definition of methodologies and techniques to assess and treat the privacy risk associated to LBS solutions. There exist many and various challenges in the deployment of LBS, but the need for identifying the risk related to the processing of personal data before determining the appropriate means to reduce them, is without doubt, one of the most important in the domain of LBS. Unfortunately, to date there is an absence of methodologies to adequately approach this problem. The risk assessment proposals found in standards [ISO08, NIS11] are so generic, that are really difficult to map to privacy and, even more to the domain of LBS.

Therefore, the main goal of this paper is to design a methodology to assess the risk related to the lack of privacy of LBS. The rest of this paper is structured as follows. Section 2 shows the lack of techniques and guidelines to assess the privacy risk on LBS. Section 3 introduces our privacy risk assess methodology for LBS. Section 4 presents a case study based on dynamic carpooling to show the usefulness of our methodology. Finally, Section 6 closes the paper.

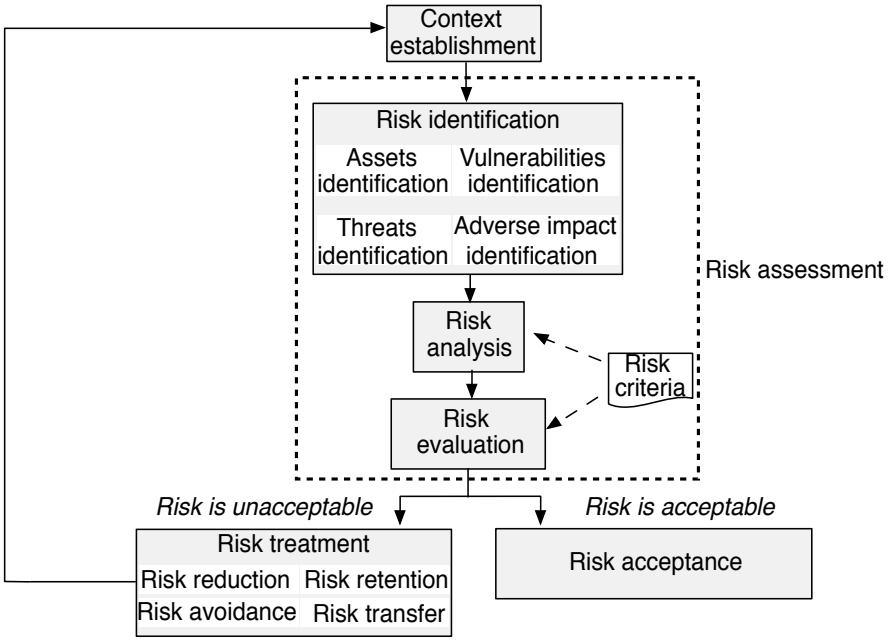
2 Related Work

The concept of risk was first introduced in safety critical systems, but is now widely used in many domains, including information technology. Indeed, users, environment and organisations could be faced to the harm induced by the use of a new technology.

2.1 Standards and regulation

The generic standard ISO/IEC-Guide73 [IG09] defines the risk as the combination of the probability of an event and its consequence. This definition had to be adapted in the domain of security, where risk is defined as the "potential that a given threat will exploit vulnerabilities of an asset or group of assets and thereby cause harm to the organisation" [ISO08, Com93, 2]. In this definition, the classic notion of *probability* of an event has been replaced by "potentiality" given the difficulty to estimate such a probability. The concept of *consequence* was also refined into "harm to the organisation". The identification, analysis

FIGURE 1 – Risk management process adapted from ISO 27005 [ISO08].



and evaluation of the risk, is defined in many standards and international directives as *risk assessment* within the risk management process, as Figure 1 shows.

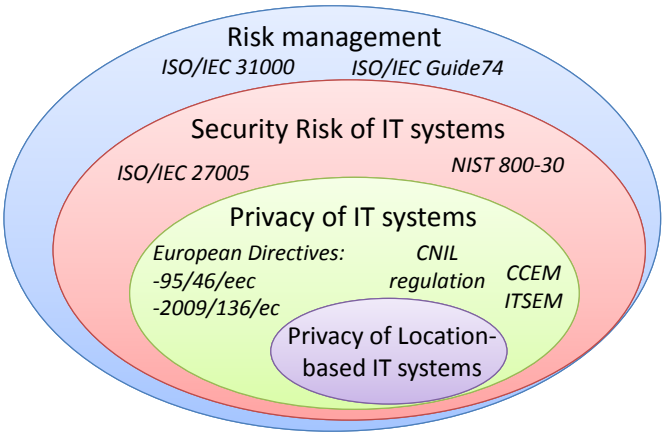
ISO standards such as [ISO08] or regulatory documents produced by NIST [NIS11], deal with security risk assessment and treatment. Unfortunately, there is no ISO/IEC standard dedicated to privacy risk assessment. In order to protect user data, the European Commission unveiled in 2012 a draft [Eur12] that will supersede the Data Protection Directive 95/46/EC[Eur95]. It is mainly about openness and obligations of organisations managing users personal data, and it does not include methods and techniques to analyse and assess the risks. A similar approach is presented in the USA Location Privacy Protection Act of 2012 (S.1233) [Con12], in order to regulate the transmission and sharing of user location data. As in the European Directives, this bill specifies the collecting entities, the collectible data and its usage, but no analysis techniques or methods are proposed, and much less in the domain of LBS.

In all those standards, the risk management process and the concepts are similar and applicable to privacy risk management. Nevertheless, as presented in Figure 1, we have simplified the process presented in [ISO08], removing some decision points for risk acceptance. The main challenge when applying this process to privacy risk management is the choice or methods for risk identification and metrics (risk attributes scales in the figure) for risk analysis. Indeed, if in security the identification can be based on experience and database for threats and attacks, it is not the case for privacy risk, and moreover for innovative location-based system.

2.2 Privacy risk models

In the last decade, we have seen a special interest in the scientific community for the development of formal methods to describe systems. However, the use of formal methods in the industrial development of security and privacy risk assessment systems is still rare. A significant barrier is that many formal languages and formal analysis techniques are unfamiliar and difficult to understand and to apply for engineers. Designers must also communicate between specialists of different domains who usually have their own language. We have seen the proposal of some component models, like the HAZOP (Hazard Operability) analysis, to identify software critical components by evaluating their risk factor. However, these methods cannot be directly applied to widely-used languages such as the Unified Modeling Language (UML) since the notion of object is too far from the notion of component they use [BRJ99].

FIGURE 2 – Risk management from high level directives to low level standards.



This trend has changed in the past years with proposals like the CORAS project [FKG+02], that introduces model-based risk assessment methodology for security-based risks using UML models. Similarly, in [Obj08], the Object Management Group (OMG) proposes a generic UML class diagram model for risk assessment, but it is very complex, and does not focus on geo-privacy issues. Despite these techniques are based on standard UML models and provide a formal method to assess risk, in general, they do not address the geo-privacy risks targeted in this paper, and because of this it is difficult to adapt the risk assessment results to this particular domain. In our prior work in [FKG13], we posed the problem of risk assessment in geo-location systems. In this paper we make a step forward to propose, to the best of our knowledge, the first methodology to quantitatively address the privacy risk associated to the lack of privacy in location-based services.

3 A privacy risk assessment methodology for LBS

The privacy risk assessment methodology introduced in this Section systematises the identification, analysis and evaluation of privacy risks for LBS. This paper more precisely focuses on showing how the use of traditional risk assessment can be coupled to a object oriented system modelling process in order to guide the designer to exhaustively consider all potential privacy risk.

3.1 Process overview and main concepts

Basically, our methodology addresses the privacy risk assessment following the framework of Figure 1. The first step, risk identification involves the identification of risk sources, events, their causes and their potential consequences. The level of risk is then estimated in the risk analysis step. For this step, we will introduce metrics as risk criteria for the privacy risk analysis. To clearly illustrate the concepts used in our methodology, Figure 3 shows the general meta-model concerning the entities and relations involved in the process. It is worth noting that our meta-model should not be seen as a closed proposal. Instead, it is a flexible abstraction gathering well-known generic security risk concepts. This meta-model can be interpreted under the prism of geo-privacy if completing it with the notions introduced in Figure 3, that focus on the identity and location assets.

As seen in Section 2, risk is generally defined as the combination of two factors : the *risk likelihood*, which refers to the frequency of occurrence of potential menaces for the system (threats) ; and the consequences that such threats may cause on victims, also referred to as *risk adverse impact*.

Assets refer to the key information used in the system. Assets can be basic or inferred (if they are obtained from basic ones). In the case of LBS, assets are mainly intangible. Among them, *identity* and *location* are the essential assets that characterise our meta-model for LBS, as introduced in Figure 4. For

FIGURE 3 – General meta-model for security risk assessment.

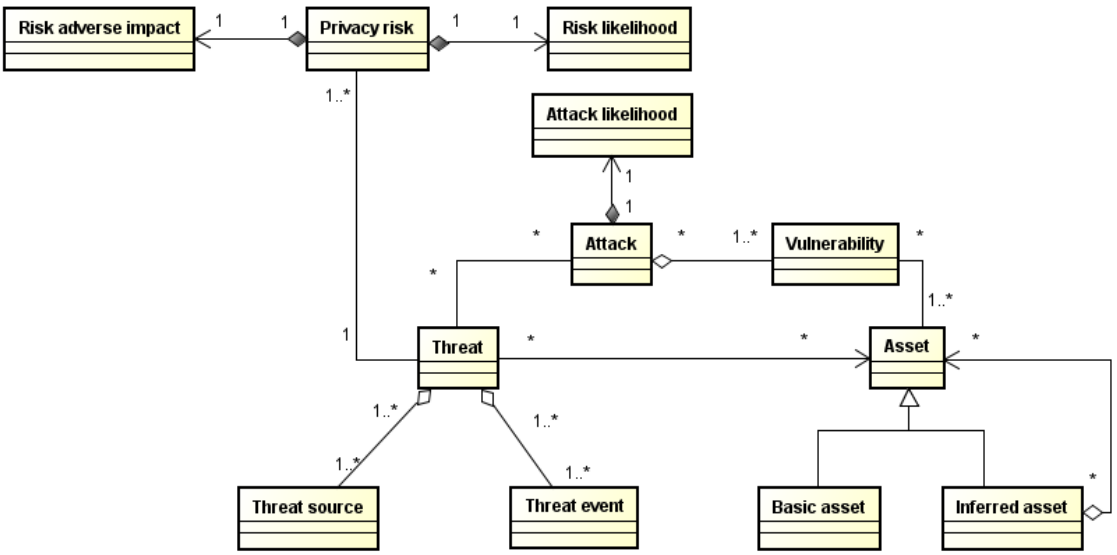
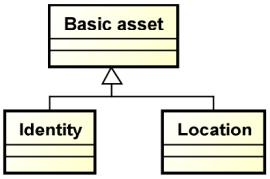


FIGURE 4 – Specific assets for geo-privacy risk assessment.



instance, the spatio-temporal data of an individual can be used to infer his movements and habits, to learn information about his centre of interests or even to detect a change from his usual behaviour. From these basic assets it is possible to obtain derived ones. It is to note that the asset value strongly depends on its precision. Fine-grained information concerning the who-where-when tuple is much more valuable than a coarse-grained one. For example, the value of knowing that a person, identified by the full name, will be at a place with a geo-temporal error of meters and seconds is higher than knowing that a person, only identified by the surname, will be potentially at a given city within an error window of two weeks.

Threats are typically characterised by the *threat sources* and the *threat events* affecting a *privacy asset*. Conversely to security, in the domain of privacy, the notion of *threat* particularly refers to the *disclosure* of assets. In most of the cases, threats are based on the potential correlation of partial informations (premises), assumed to be true, to derive more valuable information. To obtain such premises, attackers may appeal to the *usurpation* of identity, the *deception* of information or the *disruption* of services. A threat, to be realisable, needs to be instantiated in practice through attacks. It is reasonable to think that the same threat, e.g., the disclosure of the home address, can be instantiated by different attacks, e.g., by physically following the victim node, or by capturing a message containing this information. Consequently, the attack likelihood may vary depending on their level of difficulty, thus conditioning the risk likelihood associated to such a threat. To be successful, an attack needs to exploit a *vulnerability* associated to the asset. Vulnerabilities refer to inherent weaknesses in the security procedures or internal controls concerning the asset. Next sections exploit this meta-model to approach privacy risk assessment in practice.

3.2 Risk identification

Our methodology aims at covering the existing gap in privacy risk assessment by providing a simple strategy to identify such entities in the domain of LBS following a natural from-asset-to-adverse-impact approach. The goal of risk identification is to collect the information concerning the meta-model in Figure 3 in a simple but structured way, for instance, as shown in Table 1. This process is similar to the performed in the Failure Modes, Effects and Criticality Analysis (FMECA) [1]. First, it is necessary to identify the privacy assets. Among all the *Personal Identifiable Information*, learning the location of an individual is one of the greatest threat against his privacy. Essentially because starting with the location, many other private information can be derived [GKNndPC11]. For example, by composing the assets in Figure 4 it is possible to infer additional ones such as the following :

- *Social relations between individuals* by considering for instance that two individuals that are in contact during a non-negligible amount of time share some kind of social link. This information can also be derived from mobility traces by observing that certain users are in the vicinity of others.
- *Itinerary*, can be defined as a collection of locations detailed for a journey, especially a list of places to visit. An itinerary could be referred to as both physical and symbolic information containing a starting and ending point.
- *Important places*, called *Points Of Interests* (POIs), which characterise the interests of an individual. A POI may be for instance the home or place of work of an individual. Revealing the POIs of a particular individual is likely to cause a privacy breach as this data may be used to infer sensitive information such as hobbies, religious beliefs, political preferences or even potential diseases.
- *Mobility patters of an individual* such as his past, current and future locations. From the movement patterns, it is possible to deduce other informations such as the mode of transport, the age or even the lifestyle.
- *Mobility semantics* of the mobility behaviour of an individual from the knowledge of his POIs and mobility patterns. For instance, some mobility models such as *semantic trajectories* [GKNndPC11] do not only represent the evolution of the movements over time but also attach a semantic label to the places visited.
- *Linkable records of the same individual*, which can be contained in different geo-located datasets or in the same dataset, either anonymised or under different pseudonyms. For example, the association of the movements of Alice's car (contained for instance in dataset A) with the tracking of her cell phone locations (recorded in dataset B).

TABLE 1 – Overview of the risk identification table.
Risk Identification Table

Asset	Threats	Attacks	Vulnerabilities	Risk
a1	t1 on a1	at1	v1	r1
		at2	v1	
	t2 on a1	at3	v2	r2
a2	t1 on a2			

Thus, once assets fixed, reasoning about the pertinent threats related to a particular asset is easier, for example inference attacks to disclose the POIs of an individual. Then, vulnerabilities will be defined as the lack of control mechanisms enabling the potential realisation of such threats on identified assets. Finally, the users of the methodology should qualitatively determine the adverse impact on the privacy of the asset. Such an intuitive technique will guide a more systematic identification of risks on the system. An example

1. http ://en.wikipedia.org/wiki/Failure_mode,_effects,_and_criticality_analysis

of this technique will be detailed in Section 4.1 when presenting the case study. The rationale applied by this strategy flows in an intuitive way. We defend the simplicity of this approach as one of its major benefits. It enables even non-skilled users to handle the complexity of identifying privacy risk issues, thus easing the rest of the risk assessment.

3.3 Risk analysis

The risk analysis stage aims at estimating the privacy risk. As seen in our meta-model, the privacy risk of a particular threat is characterised by a risk likelihood and a risk adverse impact. The risk likelihood level (RLL) depends on the success of attacks, which is associated to an attack likelihood level (ALL). In our case, the risk likelihood will be determined by the more probable attack from all the possible attacks. Conversely, the risk adverse impact level (RAIL) is directly related to the threat, regardless which is the attack that implements it. The expression in Equation 1 shows this relation.

$$\begin{aligned} &Privacy\ Risk = f(RAIL, RLL) \\ &RLL = max(ALL_i) \mid i\ \epsilon\ Threat \end{aligned} \tag{1}$$

The first step to estimate the RLL and the RAIL involves the proposal of scales. However, given the difficulty to establish clear criteria, there is still a lack of reference scales in the literature by the time being. For example, NIST's Guide for Conducting Risk Assessment [NIS11] just provides very generic and qualitative scales which, if on one hand they are simple to understand, on the other they are very difficult to apply. Even more striking seems the case of well-known standards such as ISO/IEC 27005, which does not propose any kind of scale to quantitatively estimate such magnitudes. Alternatively, there are some tools for risk assessment such as [Ing10] which, despite not proposing scales, are able to characterise the potential occurrence of the attack through its cost, the technical ability, the noticeability, the economic benefits and the attacker satisfaction.

To reduce the gap between the real risk of geo-location-based systems, and the interpreted by evaluators, we propose two novel scales as a proof-of-concept to quantitatively measure both the RLL and the RAIL.

Each one of these factors is decomposed in different dimensions to base rating decisions in more objective criteria. On one hand, determining the number of dimensions and their possible values is very important for the feasibility of the scales. In our case we have considered 4 different dimensions at two levels each (low and high). This leads to 16 different states, which is a manageable number of combinations from the viewpoint of evaluators. According to our experience this is the most suitable configuration, since a bigger number of dimensions or levels would exponentially increase the complexity of scales. A wider discussion about this point will be introduced in Section 5. On the other hand, all dimensions should be interpreted in the same way, e.g., the higher the better, or the lower the better. By taking this decision, dimensions become comparable, which enables the establishment of common rules to interpret them. In our case, this rule is very simple and consists in assigning a 0-to-10 score to each factor depending on how many dimensions are set to low level (e.g., 4 dimensions at a low level involve a very low RAIL, but from the viewpoint of RLL this means a very high potentiality of occurrence). It is worth mentioning that this score is compliant with the proposed by NIST in [NIS11].

3.3.1 Risk Adverse Impact Level

Table 2 guides the selection of the RAIL through (i) the geographic or temporal *accuracy* of the information, that can be low if it is coarse-grained (e.g., measured in weeks or kms), or high, if it is fine-grained (e.g., measured in hours or meters) ; (ii) the *linkability* between informations, that can be low, if it is complex to relate one information with others (e.g., the location with the religious beliefs) or high, if it is easy to do so (e.g., the location with the social network) ; (iii) the *persistence* of the impact, that can be low, in case the duration of the impact is transient (e.g., disclosure of the current location), or high, if it is permanent (e.g., disclosure of home address) ; and finally (iv) the *dissemination* of the information, that can be low, if the information exposition is limited (e.g. just revealed to a small set of people), or high if it is publicly exposed (e.g., on the Internet).

TABLE 2 – Risk Adverse Impact Level (RAIL) scale. These scale is adapted to the levels proposed in [NIS11]. The quantitative value assigned depends on the amount of high and low scores obtained. E.g., 4 low scores and no high score involve a very low impact (rated as 0). This matching is possible because all the dimensions are the lower the better.

RAIL: Risk Adverse Impact Level (the lower the better)

Geographic accuracy	Linkability	Persistence	Dissemination	RAIL
Low	Low	Low	Low	0
			High	2
		High	Low	2
			High	5
	High	Low	Low	2
			High	5
		High	Low	5
			High	8
High	Low	Low	Low	2
			High	5
		High	Low	5
			High	8
	High	Low	Low	5
			High	8
		High	Low	8
			High	10

If 4 attributes are high → RAIL = 10 (very high impact)
If 3 attributes are high → RAIL = 8 (high impact)
If 2 attributes are high → RAIL = 5 (average impact)
If 1 attributes is high → RAIL = 2 (low impact)
If 0 attributes are high → RAIL = 0 (very low impact)

3.3.2 Attack Likelihood Level

Table 3 provides scales to normalise the ALL. This scale depends on (i) the *resources* required by the attack, that can be low if the computational power is reduced (e.g., a laptop is enough), or high, if much computational power is needed ; (ii) the exploit *complexity* required by the attack, that can be low, if the attack is easy to launch (e.g., a script is available on the Internet) or high, if a profound knowledge is necessary to execute the attack ; (iii) the *spatial constraints* of the attack, that can be low, if it can be performed remotely, or high, if it requires the attacker being located at some specific location (e.g., in the radio range of the victim nodes) ; and (iv) the *duration* of the necessary observation of victim nodes, that can be low if a single observation is enough, or high, if it requires many observations (e.g., for a month or even more).

3.3.3 Attack trees

If we consider the possibility of using attack trees to describe the privacy of the system, we can use scales to rate the ALL of each one of the leafs of the attack tree and then propagate their value towards the root. Thus, by following this approach, we are able to apply a systematic method to quantify the likelihood of all the attacks implementing the threat. In our case, the propagation is determined by the maximum likelihood in OR branches, while AND gates propagate the minimum value. Finally, the likelihood associated to the top event will correspond to the RLL.

Figure 5 presents a simple example to determine the risk of a well-known threat as the disclosure of the home address (as illustrated in the root of the tree in Figure 5). In this example, the threat can be implemented through three different attacks : physically following the victim to her home ; eavesdropping the applicative packets exchanged by her mobile device and filtering the home address information ; and installing a malware in either her mobile device to track her position or the server providing any service used by the victim. As computed in Figure 5, the ALL of the attacks is 5, 8 and 2 respectively. Following the

TABLE 3 – Attack Likelihood Level (ALL) scale. These scales are adapted to the levels proposed in [NIS11]. The quantitative value assigned depends on the amount of high and low scores obtained. E.g., 4 low scores and no high score involve a very high rate of occurrence. This matching is possible because all the dimensions are the higher the better.

ALL: Attack likelihood Level (the lower the better)

Resources requirements	Complexity	Spatial constraints	Duration	ALL
Low	Low	Low	Low	10
			High	8
		High	Low	8
			High	5
	High	Low	Low	8
			High	5
		High	Low	5
			High	2
High	Low	Low	Low	8
			High	5
		High	Low	5
			High	2
	High	Low	Low	5
			High	2
		High	Low	2
			High	0

If 4 attributes are high → ALL = 0 (very high prob.)
If 3 attributes are high → ALL = 2 (high prob.)
If 2 attributes are high → ALL = 5 (average prob.)
If 1 attributes is high → ALL = 8 (low prob.)
If 0 attributes are high → ALL = 10 (very low prob.)

reasoning previously introduced, the RLL is bounded by the maximum ALL. In the case of this example, this value corresponds to the eavesdropping attack, which confirms this attack as the easiest one, given its higher value. Finally, the RAIL is defined considering the threat itself. In this case, being the home address such a sensitive place, it was rated with 10 points.

To estimate the privacy risk, we consider the tuple (consequence, likelihood), represented in Figure 6 by $(RLL, RAIL)$. In this matrix, each cell characterises one particular risk value, in such a way that, a set of cells could be mapped to a risk level. For the purpose of this paper we have used a 0-to-8 scale to each cell considering 3 basic privacy risk levels. Thus, low, medium and high risk gathers those arrays rated from 0 to 2, from 3 to 5 and from 6 to 8 respectively. Figure 6 shows these risk levels. According to such a matrix, the privacy risk associated to our example, $(8, 10)$, would be rated as 7 (high risk).

Figure 7 synthesises how the entities of our meta-model are used in the risk analysis stage, and the consequent relation with the risk identification table.

3.4 Risk evaluation

The risk evaluation stage is in charge of ranking privacy risks regarding specific criteria. Applying different criteria, such as the importance of the asset for the business, the reputation of users or the regulation fulfilment, may lead to different rankings. As explained in the previous Section, other criteria could be added taking into account the context. According to the selected criterion, risks, regardless if they were high, medium or low, will be characterised as acceptable, tolerable and non-acceptable following an As-Low-As-Reasonably-Practicable (ALARP) [Mel01] strategy. The risks characterised as non-acceptable will be prioritised for their treatment. This stage is out the scope of this paper, nevertheless a short discussion is presented for the case study in Section 4.3.

FIGURE 5 – Example of attack trees.

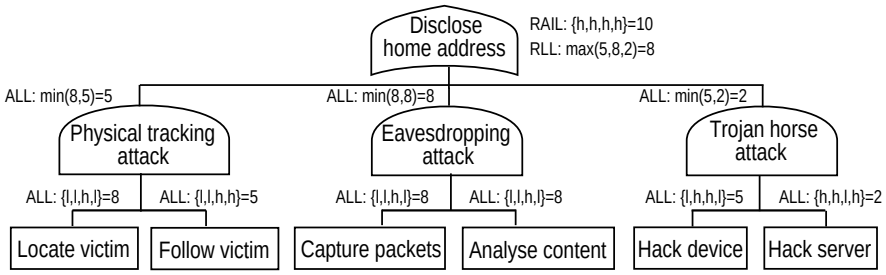
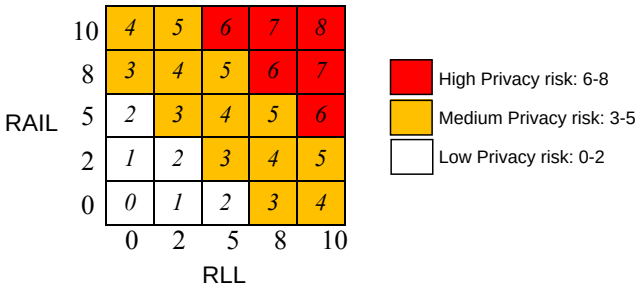


FIGURE 6 – Risk matrix considered in our methodology.



4 Case study : a dynamic carpooling application

This section shows the complete cycle of our privacy risk assessment methodology through a dynamic carpooling case study. Carpooling is an urban service in which drivers and passengers share a private car from sub-urban to urban journeys or intra-urban journeys to save money while reducing the environmental pollution. Dynamic carpooling is a wireless infrastructureless version of carpooling that implements algorithms for dynamic vehicle routing problems taking into account the mobility of users. These protocols rely on intermediate users to forward carpooling requests and responses between distant users beyond one hop. Dynamic carpooling is characterised by two actors, the driver and the passenger.

Diagram in Figure 8 illustrates the exchange of packets between Driver D and Passenger P and how these users are distributed in the network. Let us assume D belongs to the area, or geo-region A, whereas P is located at geo-region B. D and P may launch an itinerary request in which they announce the origin GPS coordinates of the journey and the expected destination, as well as their nicknames and their preferences for the trip (e.g., if they prefer travelling with smokers or non-smokers, the accepted deviation in kms from their origin or destination, and so on). Such requests will be received by all the users located in the same geo-region. In this case, $N1$, which belongs to geo-regions A and B and thus receives the requests of D and P, is in charge of processing a potential matching between these users. If there is a matching, then $N1$ forwards requests towards D and P. If D and P accept the itinerary proposed, they will send an unicast itinerary acknowledgment to the other party.

4.1 Risk identification

The personal information considered in this case study is of prime importance since it concerns the mobility and location of users. In this section we have identified those (basic and inferred) assets potentially inferred from network packets, which are of special interest given the wireless (and thus open) nature of the communication channel. These assets have been listed in Table 4.

These assets can be exposed to multiple privacy threats. Yet, given the number of assets identified, the rest of the section will focus on addressing some threats related to location and POI. As far as carpooling packets are exchanged through intermediate nodes using the wireless medium, malicious adversaries in

FIGURE 7 – Overview of the relations between the risk identification table and risk analysis stage.

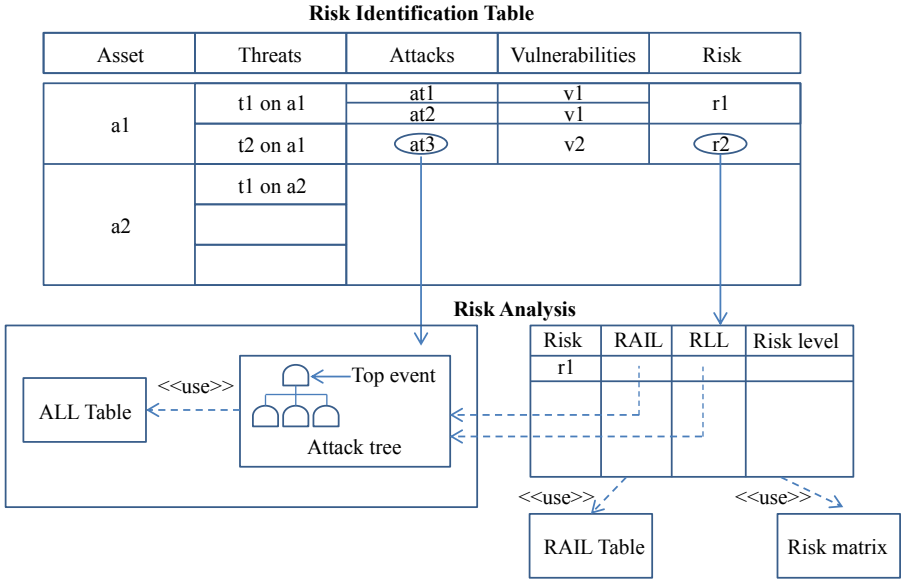
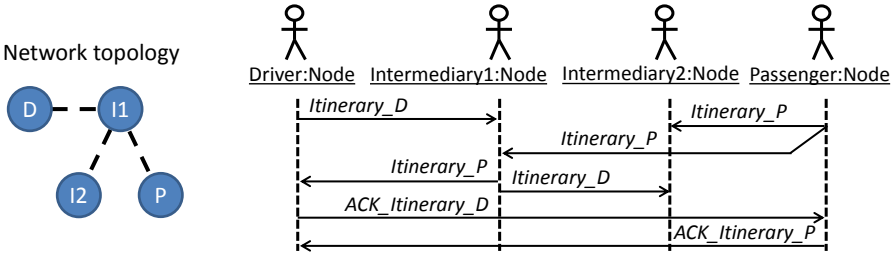


FIGURE 8 – Sequence diagram showing the general case of dynamic carpooling.



the vicinity may easily capture and analyse their content. Without the aim of begin exhaustive, the set of threats introduced hereafter relies on this principle. As these threats can be combined in order to create more sophisticated threats, they will be instantiated from the simplest to the most complex attack. Table 5 synthesises the risk identification process.

- **Eavesdropping-based identifier disclosure attack** : A malicious user could capture carpooling packets and analyse their content, including the identifier of users. Since messages are forwarded, the malicious user could eavesdrop messages from distant nodes. The absence of *confidentiality* mechanisms to protect both the *identifier of users* and the *IP or MAC address of the device* may lead adversaries to *disclose* and match the *personal identity* with the *personal identity* of users.
- **Identity spoofing attack** : Adversaries may install any malware in the user's device to remotely access her data. In consequence, adversaries may be able to carry out *non-authorised actions* on behalf of victim users, such as the negotiation of a trip they have not solicited, whose unfulfilment will penalise legitimate users' reputation for further trips.
- **Eavesdropping-based location disclosure attack** : A malicious user could capture carpooling packets and analyse their content, which includes the current location of users. Since messages are forwarded, the malicious user could eavesdrop messages from distant nodes. The absence of *confidentiality* mechanisms to protect both the exact *location* and the *IP or MAC address of the device* may lead adversaries to *disclose* and match the *personal identity* with the current or future *location* of users.
- **Triangulation-based location disclosure attack** : The complexity is higher than the eavesdropping version, since at least 3 attackers are required to estimate the location of the victim [GJK]. Attackers

TABLE 4 – Assets in the dynamic carpooling application.

Basic assets	**Description**
Identity	The user nickname identifying the carpooling packet creator and the device id related to their network identifiers (IP and MAC address).
Location	The location where a carpooling user was, is or will be.
Inferred assets	**Description**
Personal preferences	Details about the type of mate desired to share a trip with.
POI	Place with a special interest for the user, such as his work or home address.
Itinerary	Trip proposed and announced through carpooling packets by a driver or a passenger to go from an origin location to a destination location.
Interaction	Co-location in the domain of carpooling. It refers to the the itinerary shared by a driver and a passenger after a matching occurs.
Mobility semantics	Habits inferred from the user (e.g. the days he goes to the gym).

are able to infer the location of the victim after estimating and correlating one another the signal strength and the *IP or MAC address of the device* with which carpooling packets were sent. This fact makes the attack only possible to target the nodes in the vicinity of the attackers, providing an approximate version of the user's location. However, this attack is realisable even despite the use of cryptographic mechanisms to protect the *confidentiality* packet content.

– **Eavesdropping-based preferences inference attack** : Malicious users could capture carpooling packets and analyse their content, including the personal preferences specified users. Since messages are forwarded, the malicious user could eavesdrop messages from distant nodes. The absence of *confidentiality* mechanisms to protect both the *personal preferences of users* and the *IP or MAC address of the device* may lead adversaries to *disclose* and match the *personal preferences* with the *personal identity* of users.
– **POI inference attack** : Adversaries may store user's *mobility traces* to determine his POI. Mobility traces can be obtained from the instantiation of location disclosure attacks. The lack of *obfuscation* mechanisms make possible the continuous monitoring of location, thus disclosing user's POI.
– **Fake trip negotiation** : Adversaries pretending the role of intermediate nodes may *deceive* legitimate users by sending them fake messages to restart the communication protocol, thus forcing them to resend their itinerary and position. The lack of *authentication* mechanisms to protect the *location* indicated by the application may allow adversaries to compromise users' *itinerary*.
– **Eavesdropping-based itinerary inference attack** : Malicious users could capture carpooling packets and analyse their content, including the origin and destination location specified by users. Since messages are forwarded, the malicious user could eavesdrop messages from distant nodes. The absence of *confidentiality* mechanisms to protect both the *itinerary* and the *IP or MAC address of the device* may lead adversaries to *disclose* and match the *itinerary* with the *personal identity* of users.
– **Eavesdropping-based interaction disclosure attack** : Malicious users could capture and analyse the content of carpooling packets during different sessions to infer potential social relationships along time. The absence of *confidentiality* mechanisms to protect both the *itinerary* and the *IP or MAC address of the device* may lead adversaries to *disclose* and match different *personal identities* with the same *itinerary*.
– **Behaviour inference attack** : Adversaries may use the knowledge learned about user's POI to deduce data about her daily itineraries. The lack of obfuscation mechanisms to protect user's POI may be exploited by adversaries to gain a better knowledge of user's activity.

4.2 Risk analysis

As the adverse impact finally depends on the user viewpoint, we asked potential users of carpooling [2] about their effect on their private life. According to the results of the questionnaire, scaled regarding our

2. 64 people were randomly selected at LAAS-CNRS to be subjected to the experiment.

TABLE 5 – Risk identification in the carpooling application.

Assets	Threats	Attack	Vulnerabilities	Risk id
Identity	1.- Disclosure of user's identity	Eavesdropping-based identifier disclosure attack	Lack of IP/MAC address confidentiality Lack of user's identity confidentiality	R1
		Identity spoofing attack	Lack of messages authentication	
Location	2.- Disclosure of user's location	Eavesdropping-based location disclosure attack	Lack of IP/MAC address confidentiality Lack of location confidentiality	R2
		Triangulation-based location disclosure attack	Lack of IP/MAC address confidentiality	
Personal pref	3.- Disclosure of personal preferences	Eavesdropping-based preferences disclosure attack	Lack of IP/MAC address confidentiality Lack of user's preferences confidentiality	R3
POI	4.- Disclosure of user's POI	POI inference attack	Lack of IP/MAC address confidentiality Lack of obfuscation mechanisms	R4
Itinerary	5.- Disclosure of user's itinerary	Fake trip negotiation attack	Lack of messages authentication	R5
		Eavesdropping-based itinerary inference attack	Lack of IP/MAC address confidentiality Lack of user's location confidentiality	
Interaction	6.- Disclosure of user's interaction	Eavesdropping-based interaction disclosure attack	Lack of IP/MAC address confidentiality Lack of user's identity confidentiality	R6
Mobility sem	7.- Disclosure of user's mobility semantics	Behaviour inference attack	Lack of IP/MAC address confidentiality Lack of obfuscation mechanisms	R7

preliminary scales, Risks #6 and #7 were the most concerned (RAIL=Very high impact), followed by Risks #2, #3, #4 and #5 (RAIL=High impact) and Risk #1 (RAIL=Average impact). To determine the RAIL we directly applied the scale in Table 2. Figure 9 presents the attack trees of the attacks showed in Table 5. After computing the RLL of concerned threats, we obtained that Risks #4, #6 and #7 were to most probable to be carried out (RLL=Average probability), followed by Risks #1, #2 and #3 and #5 (RLL=Low probability). Thus, the privacy risk, applying the matrix in Figure 6 are illustrated in Table 6. Risks #2, #3, #5, #6, and #7 presented a high privacy risk whereas Risks #1 and #4 presented a medium level.

TABLE 6 – Risk computation in the carpooling application.

Risk id	RLL	RAIL	Risk Level
R1	Low prob. (8)	Avg. Impact (5)	Medium (5)
R2	Low prob. (8)	High Impact (8)	High (6)
R3	Low prob. (8)	High Impact (8)	High (6)
R4	Avg. prob. (5)	High Impact (8)	Medium (5)
R5	Low prob. (8)	High Impact (8)	High (6)
R6	Avg. prob. (5)	Very high Impact (10)	High (6)
R7	Avg. prob. (5)	Very high Impact (10)	High (6)

4.3 Risk evaluation

Privacy risk evaluation depends on the subjective criteria considered by the users of the methodology. To illustrate how our methodology can be customised to handle this problem, we have considered two opposite criteria. On one hand, the matrix shown in Figure 10a presents a privacy-conservative viewpoint, which could be adopted by the economic responsible of the organisation, aimed at limiting the economic investment on PETs. By playing this role, even high risks (rated up to 6 points in Figure 6 would be considered as tolerable or acceptable. In the case of dynamic carpooling, Risks #2 (R2), #3 (R3), #4 (R4), #5 (R5), #6 (R6) and #7 (R7).

On the other hand, the matrix shown in Figure 10b shows a privacy-strict viewpoint, which could be adopted by the security manager, who is aware of the importance of privacy for the success of the

FIGURE 9 – Attack trees of the case study

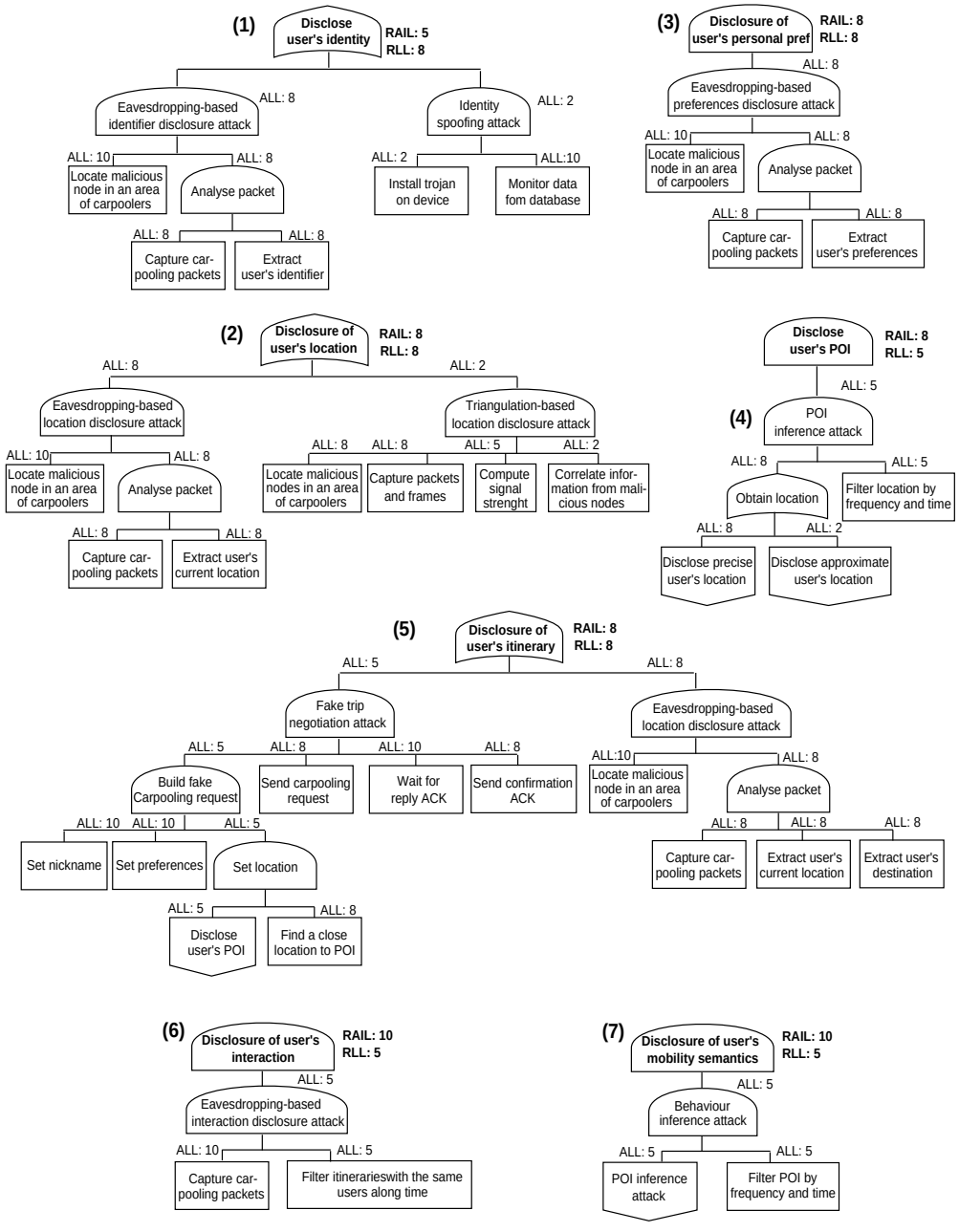
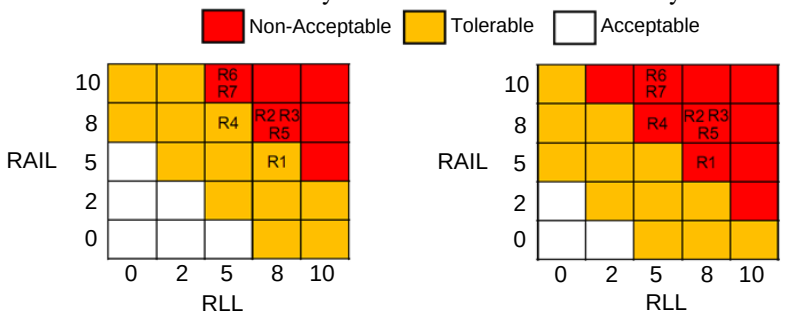


FIGURE 10 – Privacy risk evaluation of the case study.



application. Regarding this criterion, even medium risks (rated from 5 points in Figure 6 are non-acceptable, and should be prioritised for their consequent treatment. This would be the case of the risks associated to all the threats concerned. We claim that the adoption of the second viewpoint is better to safeguard the privacy of people and saves money to the organisation in a mid-term period.

5 Discussion

As seen in the paper, quantifying the privacy risk of location-based services is not an easy task given the different factors involved in this process. The present paper is a first approach towards this ambitious goal. Indeed, our goal aims at reducing the gap between complex formal scientific methods and the practical approaches solicited by practitioners. Although our methodology has been validated with a dynamic carpooling proof of concept, there are different aspects we would like to address in a near future.

Given the practical-oriented nature of our approach, we would like to evaluate its usability in real scenarios. In particular, studying the challenges (both technological and methodological) behind the implantation of this type of approaches is one of our future axis. Currently, thanks to the AMORES project, we are in contact with some companies in the domain of LBS development that are interested in its application. We think that this type of synergy is positive for the goal we pursue.

When we refer to the practical deployment of this type of approaches, it is also necessary to state the importance of the population to whom the service is addressed. Indeed, there are different contextual factors (social policy, or economic) that may condition the selection of adequate scales to measure the risk perception is targeted users. Despite the privacy risk perception is a problem that goes beyond our research scope, it is worth noting the existence of interdisciplinary studies related to psychology and sociology that address the user perception from a qualitative perspective. According to [NHP13], privacy is characterised through different attributes such as audience segregation, data sovereignty, data transience, privacy awareness, transparency or enforcement. The risk perception is studied under three dimensions : legal, technical and social. Another study shown in [Pee11] points to some key aspects that influence the privacy risk perception, such as the conditional, emotional, epistemic, functional and social values. Consequently, we argue that correctly tuning the risk scales with respect to the addressed population is of prime importance for the success of future works.

6 Conclusions

Privacy may be the greatest barrier to the long-term success of mobiquitous systems. However, despite many standards and approaches have been proposed to handle the problem of risk assessment, none, to the best of our knowledge has addressed the problem of managing the privacy risk for LBS. One of the major problems found in this paper concerns the identification of adequate information to carry out the risk assessment, as well as the way to process it. Since a lack of information may lead to obtain biased conclusions, an excess may obfuscate the decision-making process. The asset-driven strategy proposed in Section 3 provides sufficient expressiveness to guide the rest of the privacy risk assessment. Furthermore,

the risk analysis stage presented in our methodology is guided through the use of attack trees, well-known tools, to model the privacy threats, which have been expressively improved to quantify privacy risks.

The novel framework presented in this paper has been used to identify sources of risk in the lifecycle of mobiquitous solutions and find the most adequate risk trade-off between usability and privacy. Beyond this work, we are interested in studying the usefulness of our methodology to (i) guide the design PETs following a privacy-by-design approach, and (ii) compare and select (benchmark) the PETs that address the best the privacy requirements of mobiquitous systems.

Références

[BRJ99] Grady Booch, James Rumbaugh, and Ivar Jacobson. *The unified modeling language user guide*. Pearson Education India, 1999.

[2] Common Criteria (CC). Common methodology for information technology security evaluation (ccem), 2012.

[Com93] European Commision. Information technology security evaluation manual (itsem), 1993.

[Con12] Congress of the USA. S.1223 - location privacy protection act of 2012, 2012.

[DF03] J.E. Dobson and P.F. Fisher. Geoslavery. *IEEE Technology and Society Magazine*, 22(1) :47–52, 2003.

[dMHVB13] Yves-Alexandre de Montjoye, Cesar A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. Unique in the crowd : The privacy bounds of human mobility. *Scientific Reports*, 3 :–, 2013.

[Eur95] European Parliament. Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995.

[Eur12] European Commission. Proposal for a regulation of the european parliament and of the council on the protection of individuals, 2012.

[FKG+02] Rune Fredriksen, Monica Kristiansen, Bjørn Axel Gran, Ketil Stølen, Tom Arthur Opperud, and Theo Dimitrakos. The coras framework for a model-based risk management process. In *Computer Safety, Reliability and Security*, pages 94–105. Springer, 2002.

[FKG13] Jesus Friginal, Marc-Olivier Killijian, and Jeremie Guiochet. Towards a privacy risk assessment methodology for location-based systems. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Systems (Mobiquitous)*, 2013.

[GJK] Youngjune Gwon, Ravi Jain, and Toshiro Kawahara. Robust indoor location estimation of stationary and mobile users. In *INFOCOM 2004. Twenty-third Annual Conference of the IEEE Computer and Communications Societies*, pages 1032–1043 vol.2, 2004.

[GKNndPC11] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Show me how you move and I will tell you who you are. *Trans. on Data Privacy*, 4(2) :103–126, 2011.

[IG09] ISO/IEC-Guide73. Risk management - Vocabulary - Guidelines for use in standards. International Organization for Standardization, 2009.

[Ing10] Terry Ingoldsby. Attack tree-based threat risk analysis. *Amenaza Technologies Limited*, 2010.

[ISO08] ISO27005. Information technology - security techniques - information security risk management. International Standard Organisation, 2008.

[MCA06] 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.

[Mel01] Robert E Melchers. On the ALARP approach to risk management. *Reliability Engineering & System Safety*, 71(2) :201–208, 2001.

[NHP13] Michael Netter, Sebastian Herbst, and Günther Pernul. Interdisciplinary impact analysis of privacy in social networks. In *Security and Privacy in Social Networks*, pages 7–26. Springer, 2013.

[NIS11] NIST800-30. Information security, guide for conducting risk assessments. U.S. Department of Commerce, NIST), 2011.

[Obj08] Object Management Group (OMG). Uml profile for modeling quality of service and fault tolerance characteristics and mechanisms specification, 2008.

[Pee11] Loo Geok Pee. Attenuating perceived privacy risk of location-based mobile services. In *Proceedings of the European Colloid Interface Society 2011*, 2011.