Università degli Studi di Milano

Dipartimento di Informatica e Comunicazione

Rapporto tecnico

R38-11

Third-party positioning services: novel challenges for location privacy in LBS

Maria Luisa Damiani, Pierluigi Perri, Can Yizdizli

Marzo 2011

Third party positioning services: novel challenges for location privacy in LBS

Maria Luisa Damiani*, **Pierluigi Perri***, **Can Yildizli****

*Universita degli Studi di Milano (I), **Sabanci University, Istanbul (TR)

E-mail: damiani@dico.unimi.it,pierluigi.perri@unimi.it,canyildizli@sabanciuniv.edu

Abstract. A common assumption in the research community working on location privacy in location-based services (LBS) is that the location sources are trusted. In this paper we present a different perspective. We argue that, because of the deployment of wifi-based/hybrid positioning techniques and web-based LBSs, the user's location is increasingly computed by third-party location providers which may be not fully trusted. This change of perspective challenges the effectiveness of current location privacy-preserving techniques. To support this thesis we present an empirical investigation of the privacy issues raised by web-based LBSs. Moreover, following a holistic approach, we present the problem from three different and complementary angles, i.e., technical, user-based, and legal. The overall picture suggests a novel direction of research.

Keywords. Location-based services, location privacy, privacy standards

1 Introduction

Research on location privacy in LBS is strongly influenced by the evolution of the positioning technologies. For example, the assumption that has driven the research in this area over the last decade, since the seminal work of Gruteser et al. [9], is that the location source (*location provider*, hereinafter) is trusted. Typically, the position is acquired by a GPS-equipped client. Gruteser el at. [9] formulate the research problem in these terms: how to prevent the accumulation of identifiable location information in the systems of untrusted LBS providers.

Since this very first approach, research on location privacy has been developing along various directions [5]. Different taxonomies have been proposed to classify privacy protection techniques in LBS, ranging from fine-grained classifications focused on narrow domains such as [13, 1, 2], to coarse-grained classifications embracing a large spectrum of approaches such as [7, 14]. For example, Jensen et al. [13] present a detailed classification of solutions in client-server architectures, focusing on the techniques which target location privacy as opposed to identity privacy, and which apply to snapshot queries based on the users location. Krumm et al. [14] survey privacy threats which reconstruct user's identities and additional personal information from the traces of pseudo-anonymized individuals. Moreover, over the last years, LBSs have evolved from simple search-based services, e.g,. *where is the closest restaurant*, to location sharing applications, e.g., *where are my friends*. Location sharing applications introduce additional privacy requirements, for example, how to degrade the quality of the position information so as to prevent the localization of users in sensitive places, e.g., hospitals [3, 5]. Yet, in spite of the diversity of privacy-preserving methods, all of these techniques share the initial assumption, i.e., that the location provider is trusted.

In this paper we present a different perspective. We argue that in the general case the location providers, which are obviously aware of the user's position, are not fully-trusted and thus can disclose position data without the explicit user's consent. Moreover, the existing privacy-preserving techniques provide little support against this kind of privacy leak. We bring two arguments to support this thesis. Both of them are grounded on the on-going technological trend:

- Positioning technologies other than GPS are becoming increasingly popular, in particular public wifi-based positioning systems (WPS). WPS offers exciting opportunities because the position can be determined both indoor and outdoor, at an accuracy that is sufficient in most applications. Since people spend most of their time indoor, it is evident the advantage of using this technology that does not require a dedicated infrastructure, is cheap, and is commonly available on a variety of devices. The idea behind this technique is to compute the position by matching contextual information reporting the beacons (i.e., access points) nearby the device against the known location of existing networking infrastructure. Public wifi-based positioning is often used in combination with other positioning methods, such as GPS and IP address based geo-location so as to ensure a broad coverage of the territory across urban and rural areas. Currently, these so-called *hybrid* positioning services are provided by third-party location providers.

- LBSs are rapidly evolving in the direction of web applications in which services are accessed through *geo-enabled* web browsers, i.e., browsers extended with the capability of localizing the hosting device. Market studies [1] forecast that 1.7 billion geo-enabled browsers will be in use worldwide by 2016. Geo-location services can be easily added to webpages through simple scripts which are then translated into position queries forwarded to a third-party location provider. We refer to the LBSs relying on this architecture as *web-based LBSs*. Note that in such a case, both the LBS provider and the location providers can be not fully trusted that is what we wanted to emphasize.

Figure 1 shows the three components of the web-based LBS architectures: besides the client, the LBS provider and the location provider. The location provider computes the position based on the contextual information (*context*) sent by the mobile device (*client*). We call this architecture, WPS-centric[2] as opposed to the conventional GPS-centric architectures.

1.1 Paper contribution

This paper presents an empirical study on privacy issues in emerging WPS-centric architectures. In this work we want to substantiate our thesis according to which the on-going technological trend raises novel privacy concerns. Following a holistic approach, we examine the problem from three different and complementary perspectives:

[1] http://www.abiresearch.com/press/
3548-Widespread+Adoption+of+Geo-browsers+Will+Drive+Location+as+a+Key+Enabler+of+Mobile+Services

[2] We use this term for any architecture relying on a third party positioning service

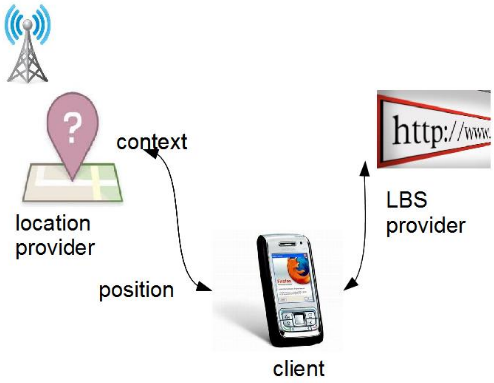


Figure 1: Web-based LBS architecture

- Technical perspective. We analyze the features of novel LBS architectures, focusing in particular on the characteristics of WPS and the emerging standard for geo-enabled browsing
- User perspective. We present a hands-on experience of position data collection conducted by tracking the visitors of geo-enabled web sites. The purpose is to gain insight into the opportunities and privacy risks raised by geo-enabled browsers
- Legal perspective. We analyze the privacy policies of two major LBS and location providers in the light of the European legislation

The paper is organized accordingly: Section 2 introduces the wifi-based positioning technology and third party positioning services. Section 3 introduces emerging standards for geo-enabled browsing and presents two experiments of users' position data collection through web-based LBSs. Section 4 presents the legal perspective. Finally in the conclusive Section we discuss a possible direction of research towards the protection of privacy against third party location providers and report conclusive remarks.

2 Public wifi-based positioning systems

2.1 Preliminaries

Wireless LAN. The wireless LAN (WLAN or wifi) technology is a family of protocols for data transmission based on the IEEE 802.11 specification [21]. The most popular among those standards, i.e, 802.11b/g operate in the 2.4GHz band. This band is free licensed and is used by a variety of low power devices, e.g., cordless phones. At this frequency, the signals propagating through the air are affected by noise and lose strength while encountering physical obstacles, such as walls and human bodies. The WLAN infrastructure consists of a set of wireless access points (APs) connected to a cabled LAN.

Wifi-based positioning systems. Positioning technologies alternative to GPS have been extensively investigated in the last decade [8, 20, 10, 12, 22]. Research on wifi-based positioning methods, i.e., methods relying on the WLANs infrastructure, develops along two main lines, targeting localization in indoor spaces and localization in metropolitan environments, respectively. The PlaceLab system [15, 11] falls into this second stream and is the precursor of the WPS commercially available today. PlaceLab allows clients like notebooks and PDAs to locate themselves by listening for radio beacons such as 802.11 APs, GSM cell phone towers, and fixed Bluetooth devices that already exist in the environment. These beacons all have unique IDs, for example, a MAC address. Clients compute their own location by hearing one or more IDs, looking up the associated beacons positions in a local database, and estimating their own position referenced to the beacons positions [15]. The authors of PlaceLab deliberately chose not to rely on any external structure and processing that could reveal user location [11]. In particular the beacons position database or part of it is kept on the client. The authors state that the concern for privacy has been one of the principles driving the system design [11].

Third party positioning services. The commercial deployment of the wifi positioning technology, initiated with Skyhook Wireless[3] is characterized by solutions generally relying on centralized architectures. Typically the beacons' position database is managed by a third party location provider while the computation of the position can be either performed by the location provider or by the client based on the content of the database. In both cases, however, the location provider is aware of the client location. Figure 2 illustrates the general architecture of a WPS, borrowed from Skyhook Wireless, that we adopt as reference architecture:

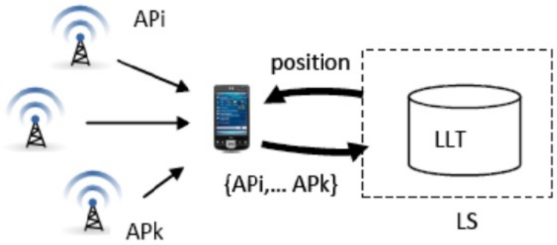


Figure 2: Reference WPS architecture

The architecture consists of a set of clients, a set of APs and a Lookup Table (LLT). The LLT is the beacon database containing the information relevant for the determination of the position, e.g., the APs location. The LLT is handled by the location provider (LS). At run-time:

- The client scans the WLANs and detects the APs in the vicinity along with the signal strength. Each AP transmits frames carrying the MAC address (i.e., the AP identifier) and additional information such as the name of the network (SSID). Hereinafter we refer to this information as the client *context*.
- Thus the client transmits the context to the location provider. The context is thus matched against the LLT, using some matching criteria. If the matching is successful, then the information associated with the APs is used to determine the position of the client.

[3] http://www.skyhookwireless.com/

```
{ "version": "1.1.0",
  "host": "maps.google.com",
........................
  "wifi_towers": [  {
      "mac_address": "01-23-45-67-89-ab",
      "signal_strength": 8,
      "age": 0},
    { "mac_address": "01-23-45-67-89-ac",
      "signal_strength": 4,
      "age": 0} ]}
```

Figure 3: Example of transmitted data

As an example of the communication protocol between the client and the location provider, we report in Figure 3 a fragment of the data transmitted by the client requesting the position to a location provider through Google Gears, an early plug-in providing geo-location capabilities to web browsers [4]. In this example, the information which is transferred to Google Location Service (i.e., the location provider) includes the field "wifi_towers" specifying the list of observed APs, each identified by the six-bytes MAC Address, along with the signal strength and an additional attribute.

2.1.1 Assumptions

For the sake of generality, and also because the information available on the internals of the existing positioning services is very limited, we do not make any assumption on the methods which compute the user's position in WPS. Computation methods include for example triangulation, centroid, fingerprint and statistic based techniques [15, 22]. Rather we consider the positioning service as a black box that can be only analyzed through its behavior. Abstractly, we model a positioning service through the function

$$geoloc(ls, \sigma, t)$$

which yields the position computed by location provider ls based on context σ at time t or null if the position cannot be computed. We recall that the context σ is a set of APs together with the signal strength of each AP, i.e., $\sigma = \{(ap_i, ss_i)\}_{i \in [1,n]}$ where n is the number of APs in the context and ss_i the signal strength of access point ap_i. Whenever σ consists of a unique AP, then the function returns a position coinciding or close to the physical location of such AP. In practice we assume that APs can be localized. Now we present an empirical characterization of the geo-location function.

2.2 Time-varying position

The contextual information in a fixed location changes in time. That is due to the variability of the signal emitted by APs. People moving in an environment, doors opening and closing, and other changes in the environment affect the signal [21, 20]. Moreover, APs can be switched on and off, especially in domestic settings. As a consequence, at a fixed location,

[4] http://code.google.com/intl/it-IT/apis/gears/geolocation_network_protocol.html

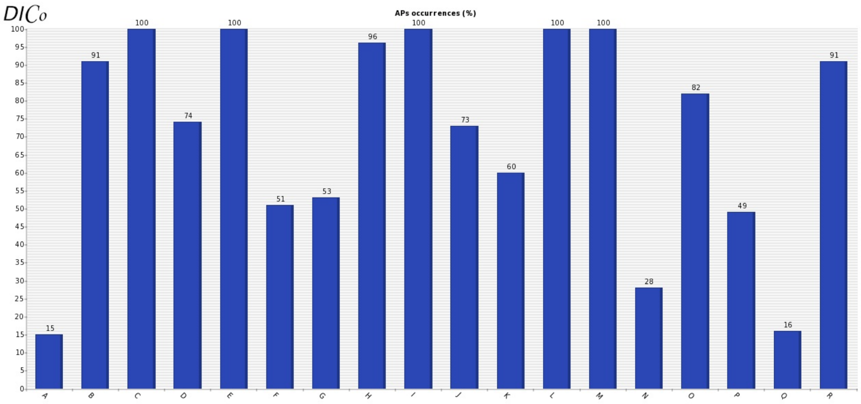


Figure 4: Frequencies of the APs heard in a fixed position in a time interval

the signal strength received from an access point varies with time. And also the number of access points covering a location varies with time [20].

We have performed an experiment to test the behavior of APs at a fixed location (i.e., the library of the University). To carry out the experiment we have developed a program which scans the WLANs nearby the client, requests the position to a third party location provider and records the association between the context and such position. The client is a wifi-enabled PC Vaio, Series Z. During this experiment, we have sampled 50 contexts at a rate of 1 sample every 50 seconds. Each context reports a set of APs. The union set of the APs present in the samples consists of 18 elements. Of these, only a subset of APs (5 out of 18) are constantly heard. The histogram in Figure 4 reports the frequencies of each AP in the time frame (the APs are labeled from A to R).

In general, if σ_1 and σ_2 are the contexts in a fixed location at time t_1 and $t_2 \neq t_1$ respectively, it holds that $\sigma_1 \neq \sigma_2$. Also the position computed by the location provider in a fixed location is time-varying, thus in general: $geoloc(ls, \sigma_1, t_1) \neq geoloc(ls, \sigma_2, t_2)$. Figure 5 shows the distribution of the positions (a subset) returned by the location provider during the experiment. In this case the accuracy of the position returned by the location provider varies between 20 and 45 meters.

2.3 The quality of the LLT data

The content of the LLT affects the accuracy and reliability of the position information. For example, if some APs among those reported in the context sent by the client, are not registered in the database, it may happen that the position cannot be returned at the best accuracy, while if the coordinates of an AP are not correct then the returned position may be unreliable. The properties of LLT completeness and soundness are important to evaluate the quality of the positioning service. Yet, very limited information is commonly available on the quality of the LLT.

The LLT typically contains millions of records. Various strategies are currently adopted to populate the LLTs. For example Skyhook Wireless populates the database through a process known as *wardriving*, namely a professional team of drivers drive around using properly equipped cars and recording observed APs and cell towers along with the corresponding GPS readings. A similar technique was initially adopted by Google through the

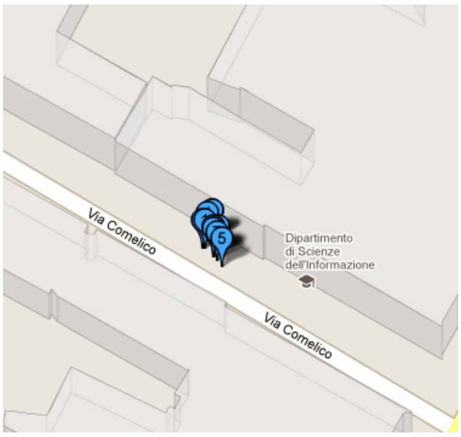


Figure 5: Position distribution in a fixed point

fleet of Streetview vehicles. Navizon applies a collaborative approach to mapping (*crowd-sourcing*), in that the position of APs and cell towers are provided by registered end users who are somehow rewarded for the collected data[5]. Another database which is fed by users on a voluntary basis is the open database Wigle [6].

The wifi network infrastructure is also extremely dynamic. i.e., APs can be easily added, removed or moved from one place to another. Following the experience of PlaceLab [15], the LLT can be updated using the contextual information sent by the clients querying their position. It has been shown however that this kind of solution paves the way to security attacks. For example, attackers can insert wrong data and/or modify existing entries [18].

Several factors thus affect the quality of data. We are not aware, however, of any indicator of the size and quality of the database apart from few statistics reported by Wigle[7].

2.3.1 Experiment

The map in Figure 6 illustrates the result of an experiment that we conducted in 2010 in a small region close to the Milan metropolitan area, using non professional wardrivers to analyze the quality of a LLT on that area. The map reports the position of both the APs registered in the LLT (yellow pinpoints) and those that have been observed during in-house wardriving (orange pinpoints). All the pinpoints are reported on a Google Earth map and each of them is labeled with the MAC Address of the corresponding AP and the SSID of the related network. We have observed that in most cases, whenever an AP is reported in the LLT, the distance from the actual position is within the tolerance threshold (i.e., within 150 meters). If so, the coordinates in the database can be considered correct. In certain cases, however, the actual position and the corresponding registered position are far from each other. An example of inconsistent positions is reported in Figure 6: the elements underlined in green refer to the same AP but there is significant distance between them. Such a discrepancy means that the coordinates reported in the database are very likely not reliable.

[5] http://www.navizon.com/

[6] http://www.wigle.net

[7] Currently, Wigle contains more then 32 millions w-ifi networks and more than 1 billion geo-referenced APs

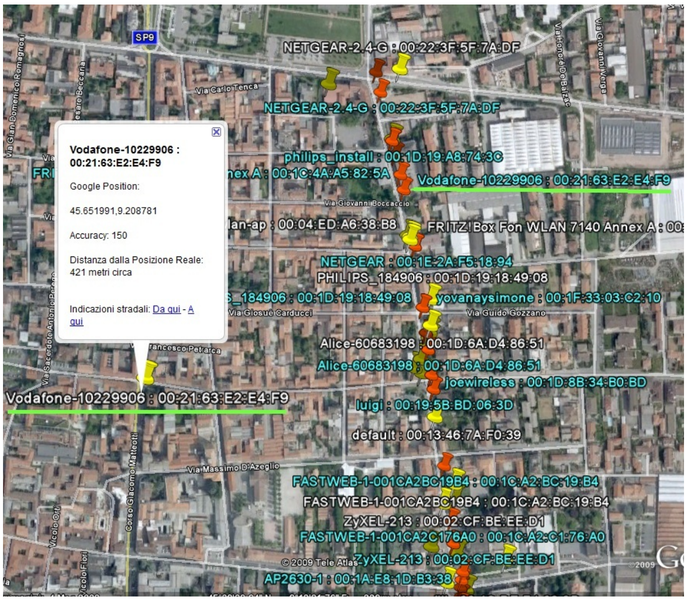

SP9
NETGEAR-2.4-G : 00:22:3F:5F:7A:DF
Via Carlo Tenca
NETGEAR-2.4-G : 00:22:3F:5F:7A:DF
Vodafone-10229906 : 00:21:63:E2:E4:F9
Google Position:
45.651991,9.208781
Accuracy: 150
Distanza dalla Posizione Reale: 421 metri circa
Indicazioni stradali: Da qui - A qui
philips_install : 00:1D:19:A8:74:3C
nex A : 00:1C:4A:A5:82:5A
Vodafone-10229906 : 00:21:63:E2:E4:F9
Via Giovanni Boccaccio
lan-ap : 00:04:ED:A6:38:B8
FRITZ!Box Fon WLAN 7140 Annex A : 00:
NETGEAR : 00:1E:2A:F5:18:94
PHILIPS_184906 : 00:1D:19:18:49:08
S_184906 : 00:1D:19:18:49:08
yovanaysimone : 00:1F:33:03:C2:10
Via Giosuè Carducci
Via Guido Gozzano
Alice-60683198 : 00:1D:6A:D4:86:51
Alice-60683198 : 00:1D:6A:D4:86:51
joewireless : 00:1D:8B:34:B0:BD
Vodafone-10229906 : 00:21:63:E2:E4:F9
luigi : 00:19:5B:BD:06:3D
default : 00:13:46:7A:F0:39
Via Massimo D'Azeglio
FASTWEB-1-001CA2BC19B4 : 00:1C:A2:BC:19:B4
FASTWEB-1-001CA2BC19B4 : 00:1C:A2:BC:19:B4
ZyXEL-213 : 00:02:CF:BE:EE:D1
FASTWEB-1-001CA2C176A0 : 00:1C:A2:C1:76:A0
ZyXEL-213 : 00:02:CF:BE:EE:D1
AP2630-1 : 00:1A:E8:1D:B3:38
© 2009 Tele Atlas
©2009
Via Gian Domenico Romagnosi
Via Cesare Beccaria
Via Sacerdote Antonio Perego
Corso Giacomo Matteotti
Vicolo Orti
Vicolo Fiori
Via Honoré De Balzac
Via Giovanni Verga
Via Francesco Petrarca

Figure 6: The APs observed on field (orange) vs. the APs reported in the database (yellow)

In synthesis, the position returned by the location provider can be incorrect, moreover the position computed by two different location providers, say ls_1 and ls_2 are likely different,i.e., $geoloc(ls_1, \sigma, t) \neq geoloc(ls_2, \sigma, t)$.

3 Web browsers with geo-location capabilities

All major web browsers have recently added the capability of geo-locating the hosting device. The diffusion of geo-enabled browsers has been pushed by emerging standards, *in primis*, the W3C Geolocation API specification [19].

W3C Geolocation API specification. This specification defines a standard set of functions which can be embedded in webpages scripts to request the position of the device hosting the browser. As an example, the key function is the *getCurrentPosition()* method: the specification prescribes that when called, the function must immediately return and then asynchronously attempt to obtain the current location of the device. If the attempt is successful, a success callback procedure is called. If the attempt fails, an error callback procedure must be invoked [19].

Note that the interface is agnostic of the underlying location information sources as well as of the geo-locating process, i.e., the way the position is obtained and by whom is completely transparent to the web page developer only depending on how those functions are interpreted by the browsers. Actually in the current implementations, the geo-location function calls are translated into queries forwarded to a third party location provider whose reference is embedded either in the browser or in the operating system.

The W3C Geolocation specification does not ignore the privacy issue. A fundamental normative requirement in the specification prescribes that the user agent must, in most cases, get the user's consent to send the device's location to a particular website before initiating a process to obtain a cached or new location [6]. Upon the request of consensus, the user can respond yes or no.

Whether this form of protection is sufficient to ensure an effective safeguard of privacy is an open issue which has not been much discussed beyond [6, 4]. In this section we want to contribute to this discussion, reporting the results of a hands-on experience of users' position data collection.

3.1 Privacy awareness: experimenting with users

The geo-location capabilities of the browsers compliant with the W3C standard can be used for different purposes. In particular, we distinguish two scenarios. In the former scenario, the user accesses a website to explicitly request a LBS. For example the users accessing the Foursquare website through a geo-enabled browser [8] deliberately share their position with the members of the geo-social network. In the second scenario, the browser solicits the user's position although the user has not requested any LBS. Since this is the case which probably raises more concerns for privacy we have performed an experiment with users. The idea is to track the users visiting certain seemingly innocuous websites. In spite of the its simplicity, this experiment allows to test a number of aspects:

- The current popularity of geo-enabled browsers
- The user's consensus (yes/no) when the position is requested by the browser
- The personal information that can be extracted from the collected location data
- The accuracy with which users are tracked

3.1.1 Experimental setting

Attribute	
a. UUID	Unique identifier associated with the client
b. IP	The ip address of the visitor
c. Position	The position of the visitor returned by the positioning service
d. Accuracy	The accuracy of the position
e. User Agent	The browser
f. Time	The time at which the web site is accessed
g. Consensus	The user's response or unsupported

Table 1: Structure of the collected data

For this experiment, we have selected two existing websites, the one located in Milan and the other in Istanbul, in order to diversify the experiments and compare the results across two different countries. The websites are accessed mostly from PCs and only for a limited percentage through smartphones or similar. In both cases, the visitors of the websites are primarily students.

The interaction scenario is as follows: a visitor accessing through a geo-enabled web browser is prompted with the question on whether he/she permits a web site, identified

[8] Using the geo-enabled version of Firefox

by an url, to monitor his/her position. The user can accept, deny, or can even decide not to answer. In any case, this choice does not have any practical and visible consequence for the user. No other information or explanation is provided to the users, beyond what is implicitly provided by the implementation of the standard geo-location interface. Therefore the url of the website monitoring the position is the only information that is shown. Transparently to the user, we gather and record for each visit the information reported in Table 1. Details are provided here below:

a. The UUID (universal unique identifier) is generated the first time a user visits the page and then stored as cookie on the user's mobile device. Actually this mechanism does not guarantee the uniqueness of visitors because users can decide to delete cookies, or it may happen that users do not allow the local storage of cookies. Another situation which my occur is that the same users access the website using different computers, for example from the university lab and at home.

c. The position is a pair of geographical coordinates. If the position cannot be determined by the location provider, either because users do not give their consent to the disclosure of position or the web browser is not geo-enabled, a coarse position is computed based on the IP address. In this case the position is the representative point of a town or of smaller areas, e.g., university campus.

d. The Accuracy field is meaningful if the browser is geo-enabled. In that case the accuracy of the position is returned by the location provider.

e. The User Agent is a string which describes various features of the web browser. The users agents which are not relevant for the analysis and which may access the web site, i.e., web crawlers, are removed from the database. The user agent is also useful to get a rough estimation of the number of different clients accessing the website as such string is a peculiar, although not exclusive, characteristic of a client.

g. We distinguish three kinds of responses: a) the browser does not support the geo-location standard. In this case the value of the field is *unsupported*. b) the user gives his/her consent: the value of the field is *yes*. c) The user denies the consensus: the value of the field is *no*.

3.1.2 Experiment 1

The first experiment has been carried out in Milan in the period October-December 2010. The website which has been geo-enabled is the homepage of one of the authors of this paper. In order to attract a sufficient number of visitors other than researchers the experiment has been carried out in the period in which the author taught an undergraduate course at the Department of Geography of the University of Milan, in the period October-December 2010. Students have a humanistic background, most of them live in the same region (e.g., Lombardy), either in Milan of in the towns in proximity of Milan. The sample data is restricted to the visits carried out from positions in the region around Milan.

Data analysis is performed over a cleansed database: the entries reporting an IP address in the domain of the University are removed to avoid multiple UUIDs for the same user as well as the entries related to user agents other than browsers. Table 2 summarizes the main features of the sample along with the geographical coordinates of the region of interest. This sample consists of 227 UUCDs while the number of visits (hits) is 940.

Sample	**#**
$region$	(8.4167, 45.9300), (10.4333, 44.9830)
$\#uucd$	227
$\#ip$	623
$\#UserAgentString$	162
$\#hits$	940

Table 2: Milan sample

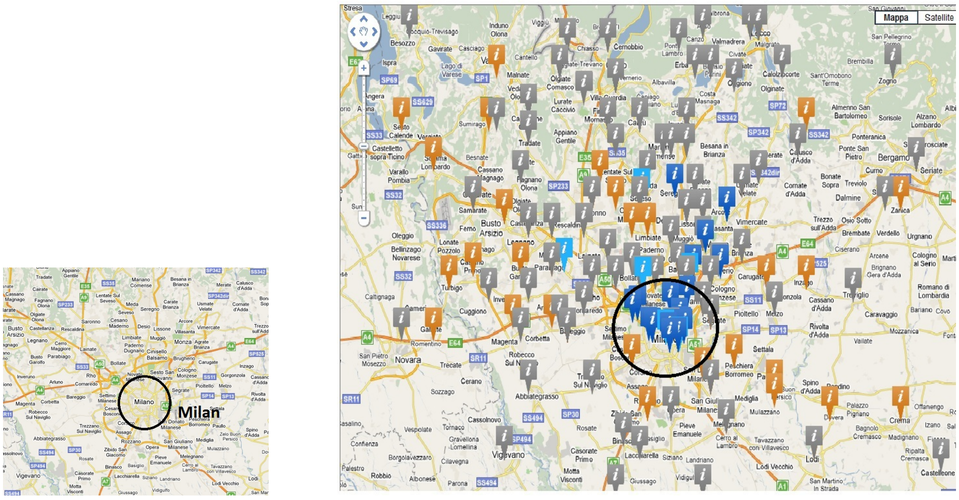
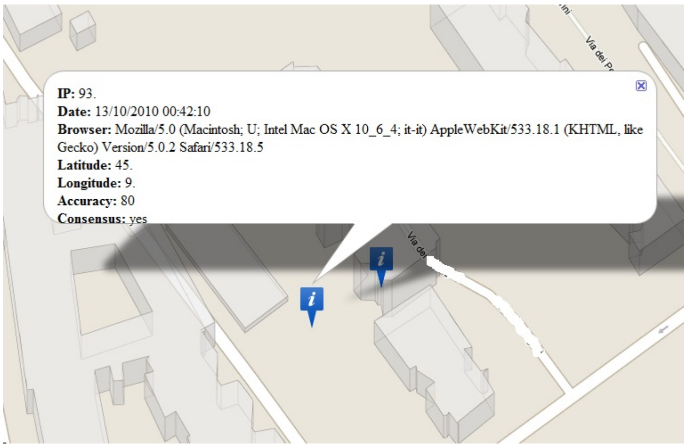


Figure 7: Experiment 1. The locations of the visitors. Color conventions: *grey* pinpoints means that the browser does not support geo-location; *orange*: the user denies the consensus to the geo-location; *blue and sky blue*: the user gives his/her consent

Qualitative considerations.

- The map in Figure 7 displays a subset of the collected data plotted onto Google maps. Each pinpoint corresponds to a visit and shows the position of the user at the time the web site has been accessed. Note that the same position can be occupied at different times. The color of the pinpoint highlights the consensus given by the most recent visitor of the web site located at that point: the gray color is used for *unsupported* consensus; orange for *no* consensus; dark blue and sky blue for *yes* consensus. In the latter case, sky blue is used for those positions which have an accuracy higher than 1000 meters. This situations occur for example when the location provider is not able to determine a position unless using an IP address. It can be noticed that the locations are scattered in the region. That is due to the fact that a significant percentage of visitors live in towns close to Milan, and it is likely that the website is accessed from home. It is also evident that the the majority of people uses a browser which does not have geo-location capabilities; there is also a significant percentage of people that deny their consensus to the position disclosure.

- In spite of the limited information stored in the database, it is relatively easy to extract personal information, such as the user's home address. As an example consider the position reported in the Google map in Figure 8. Besides the pinpoint, the Figure



IP: 93.
Date: 13/10/2010 00:42:10
Browser: Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; it-it) AppleWebKit/533.18.1 (KHTML, like Gecko) Version/5.0.2 Safari/533.18.5
Latitude: 45.
Longitude: 9.
Accuracy: 80
Consensus: yes

Figure 8: Detail of the collected data

also shows the IP address, the position, the accuracy, the browser and the consensus. For privacy reasons, we have removed the road names from the map and in part the geographical coordinates and the IP address. It can be noticed that the user has a Macintosh PC and accesses the web site at around 0:42. It is thus very likely that the user lives in the buildings in proximity of the blue marker because the position is relatively accurate (80 meters) and at that time, around midnight, the user is likely at home. This hypothesis is also supported by the fact that there is another position in proximity which is associated with the same UUID and has been detected in a different date. Moreover the positioning technique is very likely WPS. This follows from considerations on the accuracy, the type of computer and browser, the nationality. All that seems sufficient to identify the user as an individual living at that address.

Quantitative considerations. Table 3 reports the percentages for the different types of consensus (i.e., unsupported, yes, no) computed over the number of hits (% on hits) and the set of UUIDs (% on UUID). Moreover we report the relative percentage of yes/no over the number of UUCDs (% on supported). In very few cases, the same user (i.e., UUID) exhibits a different type of consensus. Table 4 reports for each browser the number of UUIDs for each type of consensus. Note that we do not report the different versions of the browsers. Normally the oldest version of a browser do not support geo-location. The exception is Microsoft Explorer (MSIE) which can be geo-enabled by certain plug-ins (e.g., GTB 6.6).

Consensus	**% on hits**	**% on UUIDs**	**% on supported**
$Unsupported$	59	63	-
Yes	4	7	19
No	37	30	81

Table 3: Users consensus

Browser	#total UUID	#unsupported	#yes	#no
IPhone Safari	1		1	
MSIE (6.0-8.0)	128	128		
MSIE(6.0-8.0 with enabled geolocation)	50			50
FireFox	15	7	5	3
Opera	1		1	
Chrome	2		2	
Macintosh Safari	18	2	8	8

Table 4: Types of browsers

3.2 Experiment 2

The second experiment has been carried out in Istanbul in February 2011. The web site which has been geo-enabled ($(http://www.confessu.org)$ is extensively used by the students of the Sabanci University to share their feelings, thoughts and confesses anonymously. The site does not require a registration so everyone can write something or read each others messages easily. In this experiment we have tried to examine to what extent the relation of trust between the user and the website depends on the name, i.e., url, of the website requesting the position. We remind that no additional information is given to users a part such url.

Who is requesting my position? When the user visits the web site, the browser pops up a message like: *<url> wants to know your location*. Normally the *url* which is notified coincides with the url of the website, meaning that the user's position is requested by the website owner. During this experiment, we use an expedient to replace the actual url with names of domains having different appeal. The expedient is to add a html *<iframe>* tag to the home page of the website. The *iframe* (inline frame) is as a frame inside a webpage which may contain another page from another website. The *iframe* tag is supported by all major browsers, moreover it is often used for malicious intents. We use two different urls. In the first case, we show the domain name "cryptovirology.org" which is likely considered as malicious. The notification is: *ge01p.cryptovirology.org wants to know your location*. In the second case, we choose a more friendly domain name, "googleservices.gmaps.me".

The experiment lasted 12 days. In spite of the short time, there is some evidence that the level of trust changes depending on the website requesting the position. The percentage of people who agree to share their position is low. Consider however that users interact anonymously with this system and thus reasonably they do not wan to reveal personal information. The great majority of users do not use geo-enabled browsers.

Sample	phase 1	phase 2
$\#UUID$	595	723
$\#ip$	419	479
$\#UserAgentString$	124	146
$\#hits$	1022	1564

Table 5: Istanbul sample

Quantitative evaluation. The sample is characterized in Table 5. The experiment consists of two phases, phase 1 and phase 2. In the former phase the url which appears in the notification is ″ cryptovirology.org″ while in the second phase is ″googleservices.gmaps.me″. Each phase has a duration of 6 days. In the first phase there are 595 UUCDs and in the second phase 723 (such discrepany is not relevant for the analysis). Table 6 reports the type of consensus in the two phases. We report the relative percentage of yes and no in each phase (column: % on support). Notice the percentage of positive consensus in the two phases. In absolute terms the percentages are low, yet in relative terms, the difference is interesting.

Consensus	**% UUIDs phase 1**	**% on supp.**	**% UUIDs phase 2**	**% on supp.**
$Unsupp.$	55.44		58	
Yes	0.33	0.7	2	4.7
No	44.22	99.3	39.98	95.3

Table 6: Users consensus

Browser	**#total UUIDs**	**#unsupported**	**#yes**	**#no**
IPhone Safari	13	0	7	6
MSIE(6.0-9.0)	595	593	0	2
MSIE(6.0-8.0 with enabled geolocation)	91	34	0	57
FireFox	181	96	3	82
Opera	5	4	0	1
Chrome	348	3	3	342
Macintosh Safari	80	17	2	61
Blackberry Safari	1	0	1	0
Nokia E72 Browser NG	4	4	0	0

Table 7: Types of browsers

Finally Table 7 reports, as in the previous experiment, the distribution of user agents for the whole period of the experiment, segmented by type of consensus. In can be noticed the richer variety of user agents, that can be in partly explained with the different background of students.

3.3 In summary

- The results of the experiment across the two countries are somehow homogeneous
- The vast majority of users use browsers which are not geo-enabled. In both cases the most popular web browser is MSIE. At the time of the paper writing, however, the geo-enabled version MSIE 9.0 has been released. Therefore the number of users of geo-enabled browsers is expected to rapidly increase
- It is relatively easy for websites to track people and identify where people live. These sites can be created in a very simple manner. It is also relatively easy to catch more user consensus by properly tuning the notification message
- The percentage of users prompt to share their position is relatively low. Note however that people is not given any information motivating the collection of position

information or any form of compensation for it, thus the percentage refers to people ready to give away their position for nothing

4 Privacy policies of LBS and location providers

After the technical and the user perspective, we turn to consider the legal dimension. In particular, we analyze two sample privacy policies from US-based companies, a geo-social network provider and a location provider respectively, in the light of the European legislation. Preliminarily we emphasize some critical aspects of the European law related to the treatment of location information. We refer the reader to [16] for an overview of privacy laws across different countries.

4.1 The location information in the European Directive

The European law through the Directive 95/46/EC (simply *Directive* hereinafter) distinguishes between *ordinary personal data* and *sensitive data*. Personal data is "any information relating to an identified or identifiable natural person...". Sensitive data are personal data which can reveal "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life". The treatment of sensitive data is usually subject to more severe restrictions by law than ordinary personal data [9]. i.e., the Legislator requests a different treatment based on the very nature of personal data. Indeed, such a strict categorization falls short when personal data have a mutable and complex nature and fall in part under the definition of ordinary personal data and in part under the definition of sensitive data.

Directive 2002/58/EC defines location data as "any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user or a Publicly available electronic communications service". The location information collected by location and LBS providers is an example of information presenting a mutable and dual nature.

When users are tracked, sensitive information can be disclosed regardless of the user's consent granted at the moment of the service request. For example, data can reveal that the user has to undergo a special daily treatment at a hospital, or that the user has requested a service designed for people with specific sexual preferences[10] Moreover, the nature of location information can change, for example it can initially represent ordinary personal data and later on become sensitive upon the occurrence of certain events.

It is therefore clear that the diffusion of LBSs calls for an integrated legal, technical and governance approach. This would create some (legal) safeguards for users and would reshape the technology so that users can actually exercise control over their own location data.

4.1.1 The on-going revision of the Directive

The need of a regulation for location data has been emphasized in a recent meeting of the European Privacy Platform group of the European Parliament [11] focusing on the ongoing

[9] Directive 95/46/EC

[10] See, for example, Grindr, a social network for gay and bisexual (http://grindr.com/Grindr_iPhone_App/Grindr_-Meet_Guys_Near_You_on_your_iPhone.html)

[11] March 16, 2011, http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/183

revision of the Directive. Such revision will be based on four pillars, reported below. Such principles will likely significantly affect location data collection practices:

- The *right to be forgotten*: the goal to ensure that individuals have really the right to withdraw consent to data processing.
- Transparency: the goal is to ensure "greater clarity" when users subscribe social networks. Moreover children must be made aware of the risks of social networks.
- Privacy by default, i.e., data protection requirements also must apply if data are processed for a purpose different from that for which they were originally collected.
- Protection regardless of data location: EU law should apply irrespective of the location of data processing and the means used by the controller to process the data. According to the Commissioner, any on-line service targeting EU consumers must comply with EU data protection law.

4.2 Analysis of sample privacy policies

A privacy policy discloses how a party gathers, uses, discloses and manages customer's data. We consider the privacy policies of two major service providers and discuss to what extent those policies comply with the guidelines of the Directive:

- *Foursquare* is the provider of a location-based social networking application[12]. Users can share their location with friends while collecting points and virtual badges. The overall goal is to allow users to bookmark information about venues they visit and get suggestions about nearby venues. The social network can be accessed through a mobile application or through a geo-enabled browser (equipped with specific plugins, i.e. FourSquareFox for Firefox).
- *Google Location Service* provides positioning services to several geo-enabled browsers (i.e., Chrome, Firefox). Our analysis focuses on the privacy practices specific to the Google Location Service (simply Google, hereinafter) that provides geo-location information to the Mozilla Firefox Geolocation Feature[13].

In this analysis, we consider the following aspects: the collected data; the granularity of the processed data; the user's consent; and data retention. Below we report and comment excerpts of the privacy policies related to these aspects.

4.2.1 Collected data

Foursquare: "automatically receives and records information on our server logs from your browser or mobile platform, including your location, IP address, cookie information, and the page you requested".
Google: "will collect, depending on the capabilities of your device, information about the wifi routers closest to you, cell ids of the cell towers closest to you, and the strength of your wifi or cell signal " and "For each request sent to our service, we also collect IP address, user agent information, and unique identifier of your client".

[12] http://foursquare.com/legal/privacy
[13] http://www.google.com/privacy/lsf.html

In both cases collected data comprise: 1) IP address, 2) cookie information or unique client identifier, 3) location. Moreover, Foursquare collects additional information about the user such as ”name, email address, phone number, birthday, Twitter and/or Facebook usernames,..., and browser information”. Google specifies the data first to process Firefox geolocation request, then ”to develop new features or products and services, or to improve the overall quality of any of Googles other products and services.”

Both services imply personal data processing. Note that Foursquare declares that ”We treat this data as non-Personal Information, except where we are required to do otherwise under applicable law”. We claim that, according to Directive 95/46/EC and 2002/58/EC, those data imply personal data processing and are subject to national privacy laws for all member States.

4.2.2 Granularity of processed data

Both policies state that data are only processed in aggregated form. In particular: ”*Foursquare* only uses this [automatically collected] data in aggregate form. We may provide aggregate information to our partners about how our customers, collectively, use our site, so that our partners may also understand how often people use their services and our Service”.
Google: ”Information collected above will be anonymized and aggregated before being used by Google to develop new features or products and services, or to improve the overall quality of any of Googles other products and services. This means that your IP address and unique identifier of your client will be stripped out before being used by any of Googles other products or features.”

We observe that in both cases the processing methods are not clearly specified. Moreover, it is well known from the research literature, that the simple removal of users’ identifiers is not enough to protect identities and in general personal information [17].

4.2.3 User’s consent and data retention

Foursquare requires the user’s consent and the acceptance of the privacy policies. Morover it allows the user to specify privacy settings. The default setting implies the sharing of informations, but several disclaimer are included in the privacy policy to make the user aware of the available privacy options.

Google uses an opt-in scheme, i.e., websites can access the location data only if the Firefox Geolocation Feature is enabled. Google limits its responsibility specifying that ”If the website is a non-Google website, we do not have control over the website or its privacy practices. Please carefully consider any websites privacy practices before consenting to share your location with that website.” and ”All requests must be sent through your internet service provider or mobile carrier network and your service provider or carrier may have access to the request. For information regarding your service providers or carriers treatment of your information, please consult their privacy policies.”

Privacy settings are friendly in Foursquare. Following the W3C specification, the privacy options available through Firefox are only to enable/disable geo-location. Moreover, *Foursquare* has also a policy about data retention, specifying that data are retained for 90 days after the user is removed while *Google* does not specify any policy.

4.2.4 In summary

The two privacy policies are to a large extent compliant with the EU privacy principles. Some aspects, however, seem not to be sufficiently clear as requested by the Directive and the national laws, in particular the data processing methods. Another aspect that is not clear regards the identification of the entities or categories of entity to whom or which the personal data may be communicated and who or which may get to know said data in their capacity as designated representative(s) in the States territory, data processor(s) or person(s) in charge of the processing is not specified, and this is an important issue for all the data subjects that want to exercise their rights.

5 Concluding remarks

Third party positioning services bring to the attention of researchers novel challenges for the protection of privacy because the position can be disclosed to untrusted location providers. We present a possible direction of research towards a solution protecting against this kind of privacy breach.

5.1 Protecting location against untrusted third parties

The research question is: can the user access a web-based LBS without communicating the position to an untrusted location provider? Consider a user willing to share his position with the website of a trusted ecologist organization, without letting the location provider know that he is at home. In this case the LBS provider is trusted while the location provider is untrusted. Note that this scenario is specular to the one assumed by Gruteser et al. [9]. We now discuss how to deal with such an issue. As a first step, we define more precisely the privacy attack.

Threat model Let $LS = \{ls_1, ...ls_n\}$ a set of location providers. Assume that the members of LS can be either trusted or untrusted, while the LBS provider is trusted.

We say that there is *location service attack* if everytime the user accesses the web-based LBS, the position is computed by an untrusted location provider i.e.:

$$\forall ls_i, \sigma, t,\ geoloc(ls_i, \sigma, t) \neq null \rightarrow ls_i \text{is untrusted}$$

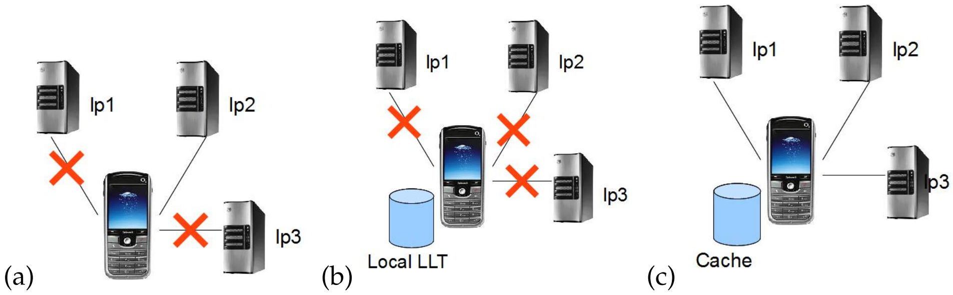


Figure 9: Protection strategies against the location service attack

Protection strategies. We envisage three different approaches that are illustrated in Figure 9. Assume a simple scenario in which a client can query any location provider in the set $LS = \{lp_1, lp_2, lp_3\}$.

- The straightforward solution is to let the user choose a trusted third party from the set LS. Figure 9(a) exemplifies the case in which the client only establishes a communication with lp_2. Obviously, this approach has little sense if the location providers are all untrusted.

- A completely different approach is to transfer the geolocation capabilities to the client like in PlaceLab [15] (Figure 9 (b)). In practice, the client maintains locally the information which is relevant for the computation of the position, i.e., a local LLT. We see two critical aspects in this approach: it is likely that this solution has a cost for the user, especially if the quality of the position information is expected to be comparable to the one provided by third party location providers. Therefore it is questionable whether this solution can reach the mass market and thus be socially relevant. The second aspect regards privacy usability, especially in relation to the need of keeping the local LLT up-to-date.

- We propose a third direction (Figure 9 (c)). The idea is not to prevent but rather to minimize the interaction with the location provider by caching a subset of the positions which have been already visited. In this way the client needs to interact with the location provider only when the requested position is not present in the cache. The deployment of this strategy raises a number of issues, for example how to select the positions to store in the cache and also how to define a suitable matching criteria taking into account that the context of a position is uncertain. A related problem is how to keep the cache aligned with the evolving WLANs infrastructure.

5.2 Conclusion

In summary, in this paper we argue that web-based LBSs offer incredible opportunities to location and LBS providers to collect huge amount of position data in a simple way. The impact of such evolution over location privacy is not clear. Also it is not clear whether the privacy mechanism currently specified by W3C is appropriate to ensure a sufficient protection. For this, the experiments with users can be of vital importance to gain insights into user's expectation on privacy.

References

[1] C. Bettini, S. Mascetti, X. S. Wang, D. Freni, and S. Jajodia. Anonymity and historical-anonymity in location-based services. In *Privacy in Location-Based Applications*, pages 1–30, 2009.

[2] P. Samarati X. S. Wang (Eds.) C. Bettini, S. Jajodia, editor. *Privacy in Location-Based Applications, State of the Art Survey*. Springer, 2009.

[3] M.L. Damiani, E. Bertino, and C. Silvestri. The PROBE Framework for the Personalized Cloaking of Private Locations. *Transactions on Data Privacy*, (3)2:123–148, 2010.

[4] M.L. Damiani and P. Perri. Privacy issues in location-aware browsing. In *In Proceedings of the 3rd ACM SIGSPATIAL Workshop on Security an Privacy in GIS and LBS*, 2010.

[5] M.L. Damiani, C. Silvestri, and E. Bertino. Fine-grained cloaking of sensitive positions in location sharing applications. IEEE Pervasive Computing (accepted for publication).

[6] N. Doty, D. Mulligan, and E. Wilde. Issues of the W3C Geolocation API. Technical report, UC Berkeley, School of Information, 2010.

[7] M Duckham and L. Kulik. Location privacy and location aware computing. In *Drummond J (ed) Dynamic & mobile GIS: investigating change in space and time. Boca Raton. CRC Press*, 2006.

[8] V. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello. Bayesian filtering for location estimation. *IEEE Pervasive Computing*, 2:24–33, 2003.

[9] M. Gruteser and D. Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *Proc. of the 1st International Conference on Mobile systems, Applications and Services*. ACM Press, 2003.

[10] J Hightower and Gaetano Boriello. Location systems for ubiquitous computing. *Computer*, 34(8), 2001.

[11] J. Hightower, A. LaMarca, and I. E. Smith. Practical Lessons from Place Lab. *IEEE Pervasive Computing*, 5:32–39, 2006.

[12] L. Hui, H. Darabi, P. Banerjee, and L. Jing. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 37:1067 – 1080, 2007.

[13] C. S. Jensen, H. Lu, and M.L. Yiu. Location Privacy Techniques in Client-Server Architectures. In *Privacy in Location-Based Applications: Research Issues and Emerging Trends*. Springer-Verlag, 2009.

[14] J. Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, (13)6:391–399, 2009.

[15] A. Lamarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F.Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilit. Place Lab: Device Positioning Using Radio Beacons in the Wild. In *Proceedings of the Third International Conference on Pervasive Computing*, 2005.

[16] M Langheinirch. *Privacy in ubiquitous computing*, chapter in Ubiquitous Computing Fundamentals, pages 95–160. CRC, 2010.

[17] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *Int. Journal on Uncertain. Fuzziness Knowl.-Based Syst.*, 10:571–588, 2002.

[18] N. Tippenhauer, K. Rasmussen, Christina C. Pöpper, and V. Srdjan. Attacks on public WLAN-based positioning systems. In *MobiSys '09: Proceedings of the 7th international conference on Mobile systems, applications, and services*. ACM, 2009.

[19] W3C. Geolocation api specification. http://dev.w3.org/geo/api/spec-source.html, 2010.

[20] M.A. Youssef, A. Agrawala, and A Udaya Shankar. WLAN location determination via clustering and probability distributions. In *Proceedings of the First International Conference on Pervasive Computing and Communications*, 2003.

[21] Kegen Yu, Ian Sharp, and Jay Guo. *Ground-based wireless positioning*. Wiley, 209.

[22] C. Yu-Chung, Y. Chawathe, A. LaMarca, and J. Krumm. Accuracy characterization for metropolitan-scale wi-fi localization. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, MobiSys '05, pages 233–245, New York, NY, USA, 2005. ACM.