Chapter 8
Privacy in Geospatial Applications and Location-Based Social Networks

Igor Bilogrevic

Abstract The use of location data has greatly benefited from the availability of location-based services, the popularity of social networks, and the accessibility of public location data sets. However, in addition to providing users with the ability to obtain accurate driving directions or the convenience of geo-tagging friends and pictures, location is also a very sensitive type of data, as attested by more than a decade of research on different aspects of privacy related to location data.

In this chapter, we focus on two domains that rely on location data as their core component: Geospatial applications (such as thematic maps and crowdsourced geo-information) and location-based social networks. We discuss the increasing relevance of geospatial applications to the current location-aware services, and we describe relevant concepts such as volunteered geographic information, geo-surveillance and how they relate to privacy. Then, we focus on a subcategory of geospatial applications, location-based social networks, and we introduce the different entities (such as users, services and providers) that are involved in such networks, and we characterize their role and interactions. We present the main privacy challenges and we discuss the approaches that have been proposed to mitigate privacy risks in location-based social networks. Finally, we conclude with a discussion of open research questions and promising directions that will contribute to improve privacy for users of location-based social networks.

8.1 Introduction

The rate at which new online data is being generated is unprecedented. It is believed that 90% of all of the online data has been produced over the past 2 years [127]. Such data is used in various domains, including healthcare, research, agriculture, logistics, urban design, energy, retailing, crime reduction and business

I. Bilogrevic (✉)
Google, Zurich, Switzerland
e-mail: ibilogrevic@google.com

A. Gkoulalas-Divanis, C. Bettini (eds.), *Handbook of Mobile Data Privacy*,
https://doi.org/10.1007/978-3-319-98161-1_8

operations [133]. In particular, location data is extremely useful for transportation, mapping, urban design, environmental monitoring and advertisement. For instance, mobility patterns of hundreds of millions of users have been mined in order to analyse the Chinese economy [55]; in another case, location data from cell-phone users, as well as buses and taxi drivers, has been used to better understand city dynamics and environmental issues [90]; similarly, location information has been mapped to crime statistics [71] and used for poverty prediction [62]. In yet another instance, location data was used for disaster relief and coordination [111, 146].

Location is one among several aspects of a person's context, such as the time, the activity, the objects or the people in proximity of a person. In order to infer the context, people use their senses. Similarly, mobile devices require sensors to determine their context, and often also communication with third-party service providers and other devices. By being aware of their context, mobile devices can provide users with a multitude of services that enrich their experience and simplify their everyday activities. For example, location awareness enables devices to provide relevant and timely driving and walking directions, or to obtain local weather forecasts. In addition to services that use location as their core functionality, more recently location data became very relevant for online social networks, by enabling users to share their locations with their social circles, by adding location information to shared media (i.e., geo-tagging) or co-presence with other people.

Location-based services are extremely popular. In the U.S., 90% of smartphone owners reported using their devices to obtain information related to their location [98]. Similarly, one of the largest (in terms of number of registered users) online social networks that uses location data has reported having surpassed one billion monthly active users [35]. In addition to being very popular among users, location data is often processed by service providers in order to enhance their services; a recent report stated that location is among the top-3 identity-related data sources used for personalization [126]. Therefore, location data is not only valuable to the users, but also to the service providers and third parties, as they frequently use it in order to drive their revenues.

In addition to being valuable, location is also a sensitive type of data [10, 79], as it can be used to reveal aspects of one's life that go beyond the location itself. Research has shown that location traces can be used to infer one's home/work places [48, 56], political affiliations [65], activities [140], interests [94] and social networks [9, 83]. Hence, being able to control the access to and flow of location data is of paramount importance for the users. Currently both Google and Facebook, two of the largest online service providers, allow their users to manage privacy settings and controls, enabling them to decide who can see their information and how it is used to personalize online services [34, 46]. For example, Google enables its users to see, correct and delete location data about them. Similarly, Facebook allows its users to decide how location check-ins and other social features (such as friend geo-tagging) work, by limiting and removing location tags [33].

In this chapter, we discuss privacy issues for two popular use-cases of location data on mobile devices: (1) geospatial applications (such as crowdsourced mapping, urban design, crisis and poverty thematic maps) and (2) location-based social

networks (such as proximity-based friend finders, online dating, social and media geotagging, as well as event planning). We begin by discussing the increasing relevance of the geospatial applications in Sect. 8.2, which have paved the way for the current location-based services. We cover topics such as crowdsourced geographic data, geo-surveillance and their relevance to privacy. Afterwards, in Sect. 8.3 we focus on a subcategory of geospatial applications, i.e., location-based social networks (LBSNs), where we discuss their different entities and their roles. For instance, users may be concerned with what other users know or can learn about them, but they can also worry about how service providers and other third parties are using their data. Next, we present the main privacy challenges and we discuss the approaches that have been proposed to mitigate the privacy risks, by surveying solutions from both the engineering field as well as the Human-Computer Interaction (HCI) domain. It is crucial to consider these related but separate aspects, as Privacy-Enhancing Technologies (PETs) are most effective when they are intuitive and bring benefits to users [10].

8.2 Privacy in Geospatial Applications

One of the most ancient instances of geospatial applications is cartography, which can be defined as the science of creating maps.[1] Although the first examples of maps were used to describe the stars rather than Earth's surface [92], modern maps are able to capture and summarize a plethora of information about the surface of our planet and its inhabitants, such as the road networks, ocean dynamics, environmental aspects related to natural disasters and thematic maps of economic indicators. For instance, road maps have been widely used to help people decide on the optimal way to reach their destinations, whereas thematic maps—which associate a specific type of information, such as poverty or crime levels, with a geographic region[2]—are routinely employed as tools to inform and guide policy and political efforts [71].

The increase in availability of different types of maps has benefited from a wider accessibility of public geographic information and geodemographic databases [25]. For example, several countries make census data publicly available to download and use.[3] In the U.S., such data contains anonymized information, at a block-level resolution, about citizens' incomes, education levels, housing and general demographics, including ethnicity, gender, age and sex.[4] In addition to census data, some countries have started releasing geo-referenced statistics related to public safety aspects, such as crime rates. In the U.S. and U.K., for instance, police

[1] http://www.merriam-webster.com/dictionary/cartography, last retrieved Dec. 4, 2016.

[2] https://www.census.gov/geo/maps-data/maps/thematic.html, last retrieved Dec. 4, 2016.

[3] http://unstats.un.org/unsd/demographic/sources/census/wphc/default.htm, last retrieved Dec. 4, 2016.

[4] http://www.census.gov/data/data-tools.html, last retrieved Dec. 4, 2016.

departments have been releasing such data on interactive websites as of 1999 and 2005 [71], respectively. In Sect. 8.2.2 we discuss in more detail the role of thematic maps and the inherent privacy issues.

Technological advances have undoubtedly helped to expand the accessibility of geo-referenced data, which has evolved in terms of both quantity and quality of the information it conveys. Currently, high-resolution satellite imagery can be accessed online for free from both governmental sources[5] and private companies such as Google,[6] Microsoft[7] and Esri.[8] With the advent of Web 2.0 and the mobile revolution of the past decade, which dramatically changed the way Internet users exchange information, interact and generate online content, the creation and curation of geographic data was no longer limited to the subject experts (such as geographers and cartographers). In fact, more and more people without a formal training in any of those fields started contributing geographic information through open access platforms [88], such as OpenStreetMap[9] and Wikimapia.[10] In Sect. 8.2.1 we discuss the benefits and disadvantages, from a privacy standpoint, of crowdsourced geospatial systems for both users and service providers.

So far, we have described how technological advances—amount of publicly-accessible data, technological advances and crowdsourced contributions—have increased both the coverage and detail of cartography in the past decades. By changing the way people interact with and search for geo-referenced data, such an evolution has also altered another important dimension for both offline and online users, which is privacy. In fact, each of the three aforementioned advances have had a distinct and yet complementary effect on the erosion of user privacy. First, the increase in the availability of geo-referenced data has potentially exposed demographic and social elements, such as gender, income and housing, to anyone with an Internet connection, anywhere in the world. In the era of big data, such abundance and availability has made it possible for researchers to develop algorithms that combine different sources of geo-referenced data to predict socio-economic, environmental and safety-related outcomes with high accuracy [62, 69, 71, 88]. Second, the increase in quality of the data that is collected (through, for example, high-resolution satellite imagery, widespread use of mobile devices and ZIP-code-level statistics) has amplified the effect on the erosion of privacy by pinpointing more accurately the spaces and places in which people live and interact. Third, if on the one hand citizen-contributed geographic information has dramatically increased the speed and coverage of geographic and sociographic data, it also added more uncertainty in the veracity of such data—especially in regions where more traditional data collection methods, such as surveys, are scarce and rare [62].

[5] http://earthexplorer.usgs.gov/, last retrieved Dec. 4, 2016.

[6] https://www.google.com/earth/, last retrieved Dec. 4, 2016.

[7] https://www.bing.com/maps/, last retrieved Dec. 4, 2016.

[8] https://www.arcgis.com/features/index.html, last retrieved Dec. 4, 2016.

[9] https://www.openstreetmap.org/, last retrieved Dec. 4, 2016.

[10] http://wikimapia.org/, last retrieved Dec. 4, 2016.

In the next subsections, we discuss privacy in geospatial applications from three different but related perspectives. First, in Sect. 8.2.1 we focus on the crowdsourcing aspect, by elaborating the ways in which such data is collected and how it could impact both the users that contributed it, as well as those it pertains to. Then, in Sect. 8.2.2 we discuss aspects related to surveillance and privacy, two elements that are increasingly relevant to users due to the increase in quantity and quality of geo-referenced data and big-data processing algorithms. In particular, we cover governmental surveillance and the privacy of socioeconomic and environmental factors, such as poverty.

8.2.1 Volunteered and Contributed Geographic Information

The Web 2.0 has made it possible for online users to generate and curate content on the Internet at an unprecedented scale. Geographic and geo-referenced data are two very popular types of data that have benefited from such a technology. Online social networks such as Facebook and Twitter have more than one billion mobile daily active users [36], and many of those users routinely share their exact location with other users of these services [15], by means of geo-tagged media content, check-ins to places and geo-referenced posts and tags (more about this in Sect. 8.3). In addition to contributing location information to online social networks, users are also voluntarily adding, updating and deleting geographic information from other types of platforms, such as online mapping ones. One notable example of such a platform is OpenStreetMap, where maps are "created by people like you and free to use under an open license."[11]

In both of these scenarios (social networks and online mapping), users are contributing geographic or geo-referenced data to a service. When users choose to add a geographic reference to a picture they post on a social network, they are aware that they are sharing location data with other users. Similarly, when a contributor on OpenStreetMap adds a new Point of Interest (POI) to a place, she or he knows that it is her or his responsibility to be as accurate and truthful as possible. In addition to such explicit choices to either attach location data or to contribute geographic information, there are more implicit ways in which users of online services are contributing geographic information, sometimes without even being aware of it. For instance, mobile apps that require access to location information are able to infer the coarse position even if users do not grant such access, simply due to the way IP addresses are shared by users or assigned by network operators [131].

Volunteered geographic information (VGI) is an expression first formulated by Goodchild [11, 45] in order to define the practice of generating geographic information by those who are not trained in geospatial data collection and analysis, and whose information may not be as accurate as those generated by official agencies.

[11] http://www.openstreetmap.org/, last retrieved Dec. 4, 2016.

More recently, geographers have started distinguishing between "volunteered" and "contributed" geographic information (CGI) [50, 66]. According to Harvey [50], one can define the two expressions in the following way:

Definition 8.1 *Volunteered geographic information*, or VGI, is crowd-sourced information with clarity about purposes and abilities to control collection and reuse. VGI refers to geographic information collected with the knowledge and explicit decision of a person.

Definition 8.2 *Contributed geographic information*, or CGI, refers to geographic information that has been collected without the immediate knowledge and explicit decision of a person using mobile technology that records location.

The difference between VGI and CGI relies in the way data is collected from the users: if it is an "opt-in" approach, then the data is volunteered, whereas if it is an "opt-out" approach, the data is contributed. Such a distinction is fundamental in order to better understand the differences in data quality and biases that could derive as a result of crowdsourced geographic data.

From a privacy standpoint, such a distinction between CGI and VGI is also very relevant. The opt-in approach of CGI makes sure that users have the choice whether or not to contribute data and that they are aware of it. Control over and awareness of data collection practices are two crucial aspects that affect the way people interact with online services [10, 117]. Usually, the higher is the offered control and transparency, the more comfortable are users with sharing information with the online platforms, especially because location data is one of the most sensitive types of personal data [10, 79]. In contrast to VGI, CGI is much less transparent when it comes to data collection, possible re-use and controls, because users may not be aware that such data is being collected at all [11]; a mobile device that is turned on and is connected to the Internet can continuously gather detailed data about the surroundings, such as radio identifiers (WiFi SSIDs, cellular antenna IDs, Bluetooth IDs), user identifiers (MAC adresses) and its position (GPS, WiFi trilateration). Based on results from such prior works in geography and privacy, Table 8.1 illustrates the differences between CGI, VGI and official geographic data curators and producers, with respect to different data and privacy properties. We define each of these properties as follows:

- *Quality*: It refers to the ability to ensure data-provenance [50]—attributes that allow one to assess the origin of the data as well as the processes used to collect and prepare it—as well as the trust in the contributors' accuracy when reporting geographic data. For example, geographic information produced by official entities is usually able to ensure both data-provenance and is assumed to be more trustworthy than data produced by an individual.
- *Coverage*: It refers to both the extent and detail contained in geographic information. For instance, the coverage provided by VGI contributors can be quite different depending on the region of the world that it pertains to. For example, regions in North America have a better coverage than those in southern Asia and Africa [84].

- *Freshness*: It refers to the update frequency of the geographic information. For instance, CGI data can be continuously collected and re-used, whereas official data relying on periodic surveys and census is usually more stale.
- *Legal liability*: It refers to the liability in case some geographic information offered by a service breaches contractual obligations or agreements, which could happen if, for example, a certain guarantee of accuracy was promised but not delivered [11].
- *Transparency*: It refers to the clear and open disclosure of data collection practices, processing and limits. For example, the presence of a privacy policy or informative content, describing the extent and use purposes of the data collection, contribute towards transparency.
- *Control*: It refers to the ability of users who engage with a service to be able to control the extent to which they are contributing information. It includes opt-in approaches, selective and granular information sharing and the ability to request information about oneself to be removed from the service. For example, opt-in approaches provide users with the choice of whether to contribute information to the service, whereas opt-out approaches usually require users to either accept all the conditions or not to use the service at all [50].
- *User benefits*: It refers to the presence of clear benefits for users, which derive from contributing geographic information to the service provider. For example, rescue operations after a natural disaster (such as the 2010 earthquake in Haiti [88]) have greatly benefited the affected population, as well as relatives, friends and organizations that were able to better monitor the evolution of the situation and to better prioritize the rescue efforts.

From Table 8.1 we notice that there is no single method that has the highest score in each of the aforementioned properties. With respect to privacy properties, the VGI method has clearly the highest aggregate score. However, it falls short in the data properties as data quality, coverage and legal liability, which are usually not satisfied. On the contrary, CGI has high score in data properties, thanks to the large number of samples that can be collected and their ubiquity. However, it falls short in the privacy properties, as the data collection methods, re-use practices and controls

Table 8.1 Properties of different methods for geographic content generation

	Data properties				Privacy properties		
Method	Quality	Coverage	Freshness	Legal liability	Transparency	Control	User benefits
VGI	−	−	+	−−	++	++	++
CGI	+	++	++	−−	−−	−−	+
Official	++	+	−−	++	+	−−	++

We assign scores on a 4-point scale from the lowest (−−) to the highest (++), reflecting the extent to which each method offers every listed property. For example, "Coverage" of VGI may be limited due to the lack of sufficient geographic data about certain regions but it may integrate environmental data collected from sensors which can enhance its value in specific cases (such as natural disasters or air quality monitoring)

are less prominent or in some cases nonexistent. In one instance, according to a CNET news report, locations of laptops, cell-phones and WiFi devices have been released on the Internet without an adequate privacy protection and unbeknownst to the users who generated it [86].

Although the modern concept of "personal privacy" has been introduced in 1890 by Warren and Brandeis [134], it is not until the early 2010s that location privacy received a significant attention in the U.S. legislation [66]. The introduction of the bills in the U.S. Congress (such as the Location Privacy Protection Act, Geolocation Privacy and Surveillance Act, Electronic Communications Privacy Act Amendments, and Online Communications and Geolocation Protection Act [100]) have prohibited actions such as the unlawful acquisition and disclosure of geo-location information to government agencies and the unlawful acquisition and disclosure of geo-location information from electronic communication media without users' consent [66, 100].

8.2.2 Geo-Surveillance and Big Data

The availability of modern technologies and large amounts of data ("big data") have undoubtedly benefited both society and individual citizens, but it has also enabled a more detailed and granular insight into their social and personal lives. On the one hand, CGI and VGI have had a positive effect on society and helped save thousands human lives [111, 146], as they enabled organizations and governments to respond in a fast way to coordinate relief efforts in cases of natural disasters, thanks to the almost real-time updates to online maps by private citizens and organizations operating both in the affected areas and outside [88]. Similarly, the availability of detailed satellite imagery and street-level views on cities and neighborhoods have enabled a better distribution of limited resources for city planners and managers, improving the living conditions of their citizens [68, 69]. On the other hand, however, they have opened new surfaces for possible threats and attacks to citizens' privacy through surveillance [22] and inference [62, 71, 72].

8.2.2.1 Geo-Surveillance and Privacy

Surveillance has always been an important instrument to achieve security and safety for authorities and governments. Nowadays, the availability of inexpensive mobile devices equipped with miniaturized sensors (such as GPS, microphone, gyroscopes, accelerometers, etc.) has enabled the collection of vast amounts of detailed measurements about the physical and social environments. For example, GPS traces or cell-tower identifiers can be used to infer one's home/work locations [31, 44]; Bluetooth and WiFi interface identifiers can be recorded and processed to infer social circles of their owners, by only relying on co-presence [9]; such information can be complemented by mining conversations recorded by mobile devices [136];

accelerometer readings on smartphones and smartwatches can be used to infer passwords and PIN codes [82, 96], whereas data related to throughput can be used to determine the most likely trajectory that a user has traveled [113].

In addition to citizen-owned devices (such as smartphones and other mobile devices), people's behavior can be monitored through more conventional surveillance means such as closed-circuit television (CCTV), red-light and thermal cameras, as well as biometric systems and RFID tags. In 2015, it was estimated that there were 245 million active CCTV cameras worldwide, which are used for purposes including traffic monitoring, crime prevention, property and home surveillance [26]. For example, judicial authorities in the U.K. have tagged over 600 adults and about 6000 juveniles with RFID chips, in order to assess compliance with bail conditions [124]. Similarly, the U.S. Department of Homeland Security (DHS) is using RFID-based documents to facilitate the entry and exit from the U.S., which can be read from up to 30 ft away.[12] Uteck [124] argues that although there is no right not to be observed, surveillance assaults human dignity and can change behavioral patterns [38, 101]. In particular, as surveillance becomes "permanent in its effects, even if it is discontinuous in its action" [38], it "disturbs the victim's daily activities, alters her routines, destroys her solitude, and often makes her feel uncomfortable" [112].

Crampton [22], a geography scholar, explores the role of geospatial information systems (GIS) in geo-surveillance, which can be defined as the surveillance of geographic activities [23]. He studies how mapping and GIS are used in recent-day surveillance and security, by broadly applying Faucault's historical method on "governmentality", which describes how people have governed themselves and others [39]. Within that framework, Crampton argues how the rationales for geo-surveillance can be traced back to the nineteenth century, when they were directly concerned with "governing (counting, measuring, and establishing norms) individuals and populations in their distributions across territories". Crampton argues that when privacy is contrasted with security, the balance points in favor of the latter in times of threat, and sometimes in favor of the former in times of peace [22]. Moreover, he also argues that opposing surveillance by appealing to privacy (or civil rights) is problematic because the latter can be defined in different ways. For instance, [22, 116] report that after the attacks of September 11, 2001, Attorney General John Ashcroft stated on National Public Radio that "we're not sacrificing civil liberties. We're securing civil liberties". Crampton also makes an additional point in his essay, where he argues that civil liberties are increased for people who are "normal" in their behavior, but they are reduced for the others. Norms, in this sense, are determined by computing statistical averages and likelihoods of behavior, both at the individual as well as the group levels. Thematic maps, which we discuss in the next section, have emerged after such behavioral norms and statistics have been established.

[12] https://www.dhs.gov/radio-frequency-identification-rfid-what-it, last retrieved Dec. 4, 2016.

8.2.2.2 Thematic Maps, Big Data and Privacy

Thematic maps are usually designed to illustrate a specific type of data (such as socioeconomic, environmental or health data) related to a geographic area and for a single purpose.[13] In contrast, reference maps usually show a multitude of data types (such as political, geographical and geologic) together on the same map [119]. For example, Fig. 8.1 shows both kinds of maps: on the top, a thematic map illustrates the poverty rates of the total U.S. population in 2014, by County [123], whereas the map on the bottom depicts a reference map of the same geographic region. In the former, the county borders serve only as visual enhancements for the poverty information the map conveys, whereas in the latter, the data related to political boundaries, geological information and demographics serves its own purpose [119].

As shown in Fig. 8.1, thematic maps can be used to convey different types of geo-referenced data, with varying degrees of privacy sensitivity for the citizens. Information related to financial information, physical safety and health is usually considered to be more sensitive than data related to generic demographics such as age and gender [10, 79]. In the late 1990s and early 2000s, authorities in the U.S. and U.K, respectively, started releasing information related to crime statistics at a regional level through online crime maps [71]. For instance, Fig. 8.2 shows an online crime map for the region of Berkeley, California, for crimes reported by the Berkeley police between Oct. 18–24th, 2016. As it can be seen, the map shows that there were a total of 136 records during the time period under consideration in that region, and it is possible to select individual records to obtain the time at which it was reported and the place where it happened. Moreover, the interface allows the users to filter by type of crime, region, time period, and to visualize aggregate charts and reports.

Kounadi et al.[71] start the discussion on privacy issues related to crime maps by describing four main issues. When exact locations are attached to crime events, (1) the victims may fear that offenders would consider them as particularly easy targets, (2) they would not want to help the authorities with the investigation as a result, (3) they would be reluctant to report another similar offense to the police and finally (4) that their address and other information could be misused [135]. One on the first attempts to assess re-identification risks as well as to outline the implications of sharing sensitive crime-related information was published by the UK's public body "Information Commissioner's Office" (ICO) in 2012 [60]. The publications of crime-related data has started as a result of a transparency program of the U.K. police, which had three policy objectives [18]: (1) To improve the credibility of crime statistics for the citizens, (2) to provide a more community-focused police service and (3) to inform, engage and empower the public to participate in crime prevention efforts. In the official ICO report, the authors tie the release of crime information to the number of households and frequency of updates, in an effort to provide anonymity for the victims, obfuscate the precise locations of the reports

[13] http://guides.lib.uw.edu/c.php?g=341594&p=2304475, last retrieved Dec. 6, 2016.

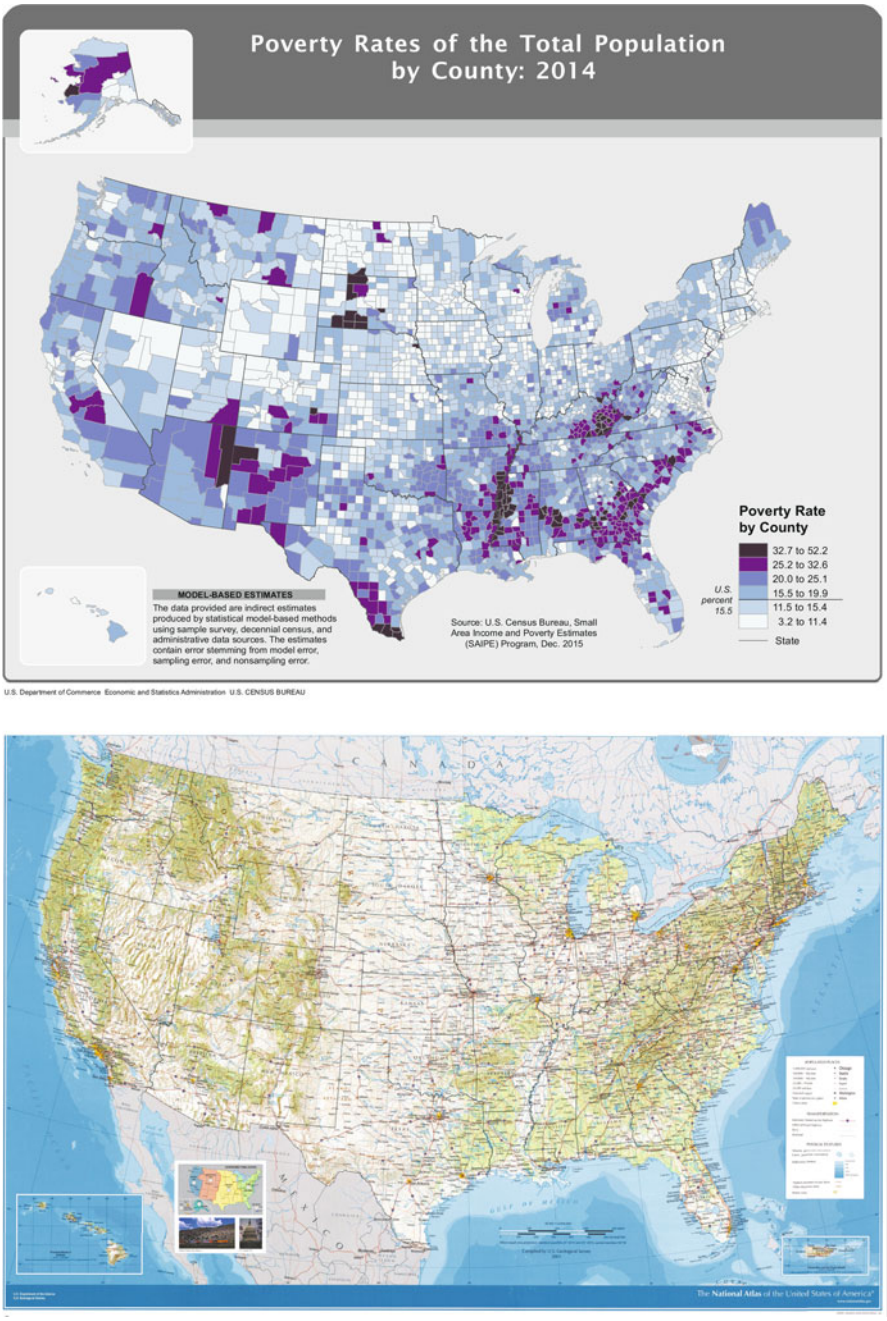

Fig. 8.1 A thematic map (top) and a reference map (bottom). The thematic maps shows the poverty rate by county in the U.S. in 2014 [123], whereas the reference map shows the U.S. territory by State, together with topographic, transportation and demographic information (images: (top) https://www.census.gov/did/www/saipe/data/statecounty/maps/iy2014/Tot_Pct_Poor2014.pdf, (bottom) https://upload.wikimedia.org/wikipedia/commons/7/7d/United_states_wall_2002_us.jpg, last retrieved Dec. 6, 2016)

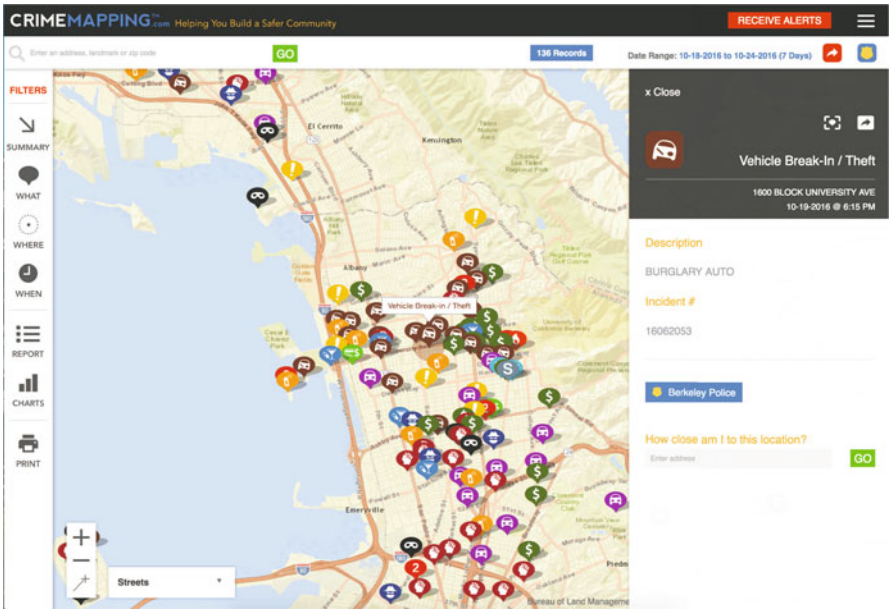

Fig. 8.2 Online crime map for the region of Berkeley, California, during a 1-week period between Oct. 18–24th, 2016. The map shows the different types of reports, such as thefts, burglaries, assaults, vandalisms, and the place where they were reported by the Berkeley Police (image shown with permission from crimemapping.com, http://www.crimemapping.com/, last retrieved Dec. 6, 2016)

and add statistical noise to them [71, 120]. Crime maps released by the U.K. police website[14] have to comply with such requirements. Although cime maps are being published, Kounadi argues that the policy objectives have not been fully achieved, in particular the one about citizen engagement and empowerment. Moreover, the participants to their study reported being more concerned with the risk implications of burglaries and violent crime statistics on maps than not, and they also expressed concerns that the released locations of burglaries could be used for commercial purposes by alarm and commercial companies (88%). However, when asked about the presence of privacy violations as a result of the release of exact burglary locations, one third of the participants that did not feel there were any violations. Such a number is certainly not insignificant, and it might indicate that for some people such information is indeed not sensitive, or it provides more benefits than risks, or that there is still misconception about the potential of geospatial tools and techniques [71].

Compared to other types of data that can be provided by any online user, the crime statistics are primarily collected by the police authorities in each country,

[14] https://www.police.uk/, last retrieved Dec 7, 2016.

and they are usually more trusted because it should be possible to verify their provenance, quality and truthfulness. Similarly, data related to socioeconomic factors such as income, education and occupation is usually collected by means of national or regional surveys by the respective governments. Often, the availability of such data is non-uniform across different countries and regions of the world, such as Africa or Asia [84]. One promising way to overcome the scarcity of official statistics about socioeconomic factors is to combine them with related data from other sources, such as satellite imagery. Jean et al.[62] have demonstrated how, by combining high-resolution satellite imagery with the survey data, they were able to explain up to 75% of the data related to economic factors such as average household expenditure or wealth. Their method relies on deep neural networks trained on both satellite images as well as existing survey data. One of the main properties that enabled authors to achieve such results is that the satellite images showed the shape and material of the rooftops, as well as the distance of the houses from the urban areas. Survey data showed that such features, which are visible in the daytime satellite images, varied roughly linearly with expenditure [62]. Moreover, the performance of the algorithm was degrading only modestly when data from one country was used to predict poverty in another country. As economic and financial data are considered to be some of the most sensitive data types, the privacy implications of fusing them with location (another sensitive data type) have only recently started to get attention by the research community. In particular, Bilogrevic and Ortlieb [10] have shown that, taken individually, location information was considered as the most personally identifying type of data, as compared to other types of data such as email address, web browsing and purchase history. However, when combined with other types of information, the combination that includes location was no longer considered as the most sensitive; a combination that included information related to online behavior, rather than offline, was considered as the most personally identifying, and thus sensitive.

Of particular concern to privacy in geospatial applications is data related to users' health conditions and their combination with location data, which can have negative effects on both the services users receive as well as on the value of their private properties [12, 14]. In many countries, medical data and records are regulated and their access and use is subject to strict access rules [4, 30]. For instance, in the U.S. the Privacy Rule in the Health Insurance Portability and Accountability Act (HIPAA) of 1996, which went into full effect in 2003 [12], applies to any individual's past, present and future data about both physical and mental health. It establishes limits to use of such data and defines which types of health data are considered "protected health information" (PHI).[15] For example, PHI includes patients' names, geographic identifiers that define a region smaller than a State (street, city, county, last three ZIP-code digits, etc.), dates (except years), telephone numbers, email addresses, vehicle identifiers, IP addresses, biometric

[15] https://www.hipaa.com/hipaa-protected-health-information-what-does-phi-include/, last retrieved Dec. 4, 2016.

data and images. In 2013, HIPAA was updated to cover additional entities, such as business associates, and it reinforced the need to disclose data breaches that previously would have been unreported [12]. Moreover, it increased penalties in case of PHI violations. Together with HIPAA, the recent legislation on location data has helped strengthen protections around two of the most sensitive data types [9], and to increase transparency in case of data breaches and leaks. In addition to HIPAA, independent institutional review boards (IRBs) are committees that have been formally created to approve, monitor, and review biomedical and behavioral research involving humans. In particular, such committees often perform a risk-benefit analysis to determine if a study should be conducted [97].

What is exempt from official IRB oversight are services that do not collect health data, and process non PHI data, such as search queries entered by online users on a search engine, in order to infer aggregate health-related trends. One such service is Google Trends,[16] which can be used to assess the popularity of different search terms over time and space. The precursor to Trends was Google Flu Trends, a service which provided flu prediction models based on patterns extracted from search queries, active between 2008 and 2015. Shortly after the launch of Google Flu Trends, the Electronic Privacy Information Center (EPIC) and the Patient Privacy Rights wrote a letter to the then Google CEO Eric Schmidt,[17] expressing concerns over the anonymity of the search queries and asking clarifications about the methods used to anonymize them. As of 2015, Google no longer publishes models directly, but it rather provides "Flu and Dengue signal data directly to partners", which include the Center for Disease Control and Prevention (CDC) [47].

So far, we have discussed how privacy concerns in geospatial applications have intensified and spread across multiple dimensions, fueled by the development of new mobile and Internet technologies, sensors, and interaction methods that allowed more and more data and people to contribute geographic information. In the next section, we focus on a more recent and very relevant subcategory of geospatial services that have received a large amount of attention and scrutiny by the privacy research community, i.e., Location-Based Social Networks (LBSNs).

8.3 Privacy in Location-Based Social Networks

Before online social networks became extremely popular over the first decade of the 2000s, Internet users relied on bulleting board systems (BBSs) instant messaging (IM) and forums in order to socialize online and exchange content [115]. Initially, online social networks such as Classmates.com [18] and Friendster[19] allowed users to

[16] https://www.google.com/trends/, last retrieved Dec 7, 2016.

[17] https://epic.org/privacy/flutrends/EPIC_ltr_FluTrends_11-08.pdf, last retrieved Dec 7, 2016.

[18] http://www.classmates.com/, last retrieved on Dec. 14, 2016.

[19] http://www.friendster.com/, last retrieved on Dec. 14, 2016. Friendster is no longer active as of Jun. 14, 2015.

search for other users they knew either by name or by affiliation to a group (such as school class or personal interests), but not much more. Later on, more recent social networks such as LinkedIn,[20] Myspace,[21] Facebook,[22] Gowalla[23] and Foursquare[24] started to integrate novel functionalities that would enable users to share more information with the service providers, and to search for and get recommendations about other people, places and activities. In particular, location APIs and location-sharing activities became more and more popular among users who were using their mobile devices to search for local content, places and people in their vicinity. By enabling users to share contextual and geographic information with the service providers, such social networks embraced the two concepts related to contributed and volunteered geographic information (CGI and VGI, respectively) discussed in the previous section: Users volunteer geographic information when they actively check-in to venues or share their locations with other users of the network, and they contribute information by simply connecting to the service from different places and devices.

There are several benefits that users enjoy if they share their location with OSNs. For instance, Foursquare users can receive location "badges" when they check-in very frequently to places and businesses. In turn, some of these businesses then provide incentives to users who have earned badges at their locations, in the form of coupons, discounts or prizes. Another popular example involves friend finder and online dating platforms. By sharing their locations, users can see other users in their proximity and engage with them, discover interesting events happening nearby and set location-based alerts that would inform them every time a given person is close to them. However, there are also downsides to location sharing. Exposing one's location renders the person more vulnerable to stalking, burglaries, physical harm and embarrassment [104]. For example, in 2010 three burglars relied on Facebook status updates to determine which houses to rob, and they managed to steal $ 200,000 worth of goods from 50 different locations [19]. A more comprehensive study conducted in 2011 showed that, based on the reports of 50 ex-burglars in England, 78% of them used Facebook, Twitter, Google Street View and Foursquare to prepare for the robberies [27]. The bridge between the online world and the physical one is clearly stated in the precise definition by Zheng of a location-based social network [144]:

> A location-based social network (LBSN) does not only mean adding a location to an existing social network so that people in the social structure can share location-embedded information, but also consists of the new social structure made up of individuals connected

[20] https://www.linkedin.com/.

[21] https://myspace.com/, last retrieved Dec. 14, 2016.

[22] https://www.facebook.com/, last retrieved Dec. 14, 2016.

[23] http://mashable.com/2012/03/11/gowalla-shuts-down/#sBOot7U3xSqf, last retrieved Dec. 14, 2016. Gowalla is no longer active as of 2012.

[24] https://foursquare.com/, last retrieved Dec 14, 2016.

by the interdependency derived from their locations in the physical world as well as their location-tagged media content, such as photos, video, and texts.

When interacting on LBSNs, users often face the question of how much location information to attach to the content they post, concerned with the possible privacy implications of their acts. While it is true that the platforms are usually designed to facilitate the sharing of geo-referenced content [114], users have very different attitudes and behavior towards sharing data online [3, 78, 118, 139]. For instance, it has been observed that, although users state they worry about the privacy of their data, they often reveal personal information on social networks [125]. The discrepancy between attitudes and behavior in the privacy domain was termed as "privacy paradox" by Barnes in 2006 [5], and is still relevant today [28]. On the one hand, some researchers argue that one way to re-conciliate attitudes with behaviors would be through the availability of better sharing controls and notices [6, 130]. On the other hand, however, some scholars believe that, although a necessary condition, better controls and notices have a limited effect on the information disclosure behavior on social networks [2].

Attitudes and behaviors aside, measuring privacy remains an open research topic. As opposed to network performance metrics such as throughput, latency, and error rate, metrics for privacy are highly dependent on the specific application and context being considered [7, 24, 54, 132]. Scholars from both the legal domain as well as engineering have attempted to classify and create taxonomies for the different ways in which privacy could be measured. For instance, Herrman [54] focused on the regulatory issues regarding compliance, operational resilience and returns on investments, whereas Wagner and Eckhoff [132] propose and categorize over 80 different privacy metrics for quantifying the privacy protection provided by privacy-enhancing technologies (PETs). In this section, we discuss privacy metrics that are directly related to the specific context of LBSNs and the privacy protection techniques that are used. More details about each of these metrics can be found in the respective paper, article or book.

In the remainder of this section, we first introduce the generic architecture of a LBSN. Next, we discuss privacy threats and protection mechanisms in five main categories: Location, absence, co-location, identity and demographics, and activity. We conclude the section with a discussion of open research challenges for privacy in LBSNs.

8.3.1 Architecture of Location-Based Social Networks

LBSNs inherit most of the standard architectural components from the traditional online social networks, which include entities (such as people and organizations) and resources (such as media or textual content), and relationships between them. Additionally, location-related information (such as location updates from users, check-ins and geotags) can be attached to both entities and resources [17, 129]. The

service provider has a central role in enabling users to connect with each other and the other entities that have an account. In order to join the LBSN, users and other entities register with the service provider, which requires them to provide some personal information such as name and email address [17]. Once the registration is successful, usually after verifying the provided email address, users and other entities can start interacting with each other and post content on the platform. Social ties and group memberships are established by asking other users and groups to join their social circles. In addition to explicitly joining social circles of other entities, often users can also opt to simply stay up-to-date with other users' updates and public posts, by means of a follower-followee relationship model spearheaded by Twitter.

Figure 8.3 shows a generic architecture of a LBSN, with a particular emphasis on the location-related aspects. In the diagram, we can see that all registered entities (people and organizations) can provide location-related information to the LBSN. For instance, people can share their current location by means of a location update and by geotagging resources such as pictures, posts, status updates and other users. Similarly, organizations can geotag resources and other organizations (either directly or through a hashtag coupled to a geotagged post). Users can obtain their current location either locally, by relying on the GPS sensor on their devices, or

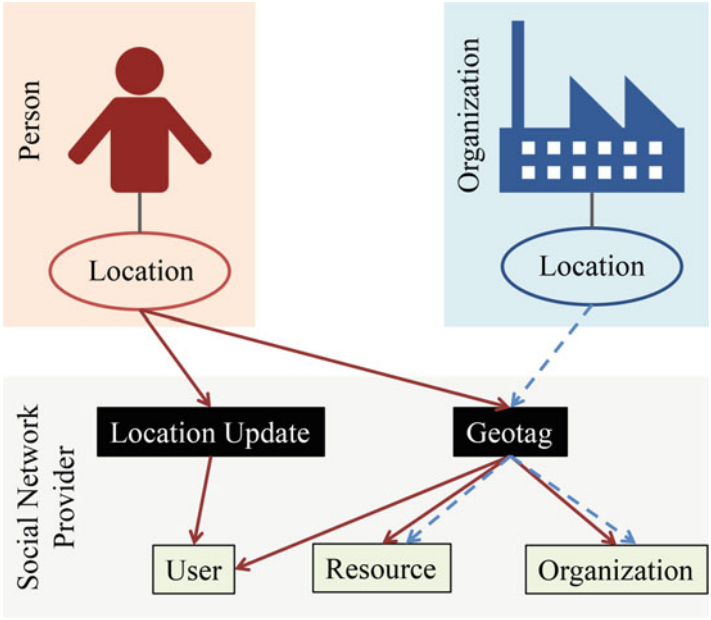

Fig. 8.3 System architecture of a generic Location-Based Social Network. The links represent possible ways by which location data can be attached to the content posted by either people or organizations (adapted from [129]). The solid and dashed lines correspond to the actions that people and organizations can perform, respectively

remotely by providing third-party services information about their current location context (such as the signal strengths and identifiers of nearby WiFi access points and cell-towers) [20, 141]).

In order to better classify the different types of LBSNs, Zheng defined three categories that capture the three main goals of a LBSN [104, 144]:

1. *Geotagged-media-based*: Service providers in this category allow users to attach location information to the content they share, such as text, pictures, videos and other types of media. For example, Twitter, Instagram, Facebook, Periscope all allow users to geo-tag content they post.
2. *Point-location-driven*: Services belonging to this category allow their users to share their current, real-time location, in order to enable a better convergence between physical and online presence, also by enabling users to discover the presence of friends (on the social network) that are in physical proximity. Moreover, such services allow users to share their experience about a certain place with other users by means of "tips" or reviews. Foursquare is an example of such a service.
3. *Trajectory-centric*: A trajectory-centric service enables users to not only share their punctual locations at different places, but also to share the route that connects them. Recently, such services have become increasingly popular, thanks to the availability of inexpensive activity tracking devices, such as Fitbit,[25] and to the increase in the type and number of sensors on mobile devices, which enable users to share their physical activity with other users, to engage in virtual competitions and to obtain virtual badges for completing activities with a certain performance [37].

Many of the popular LBSNs belong to either one or multiple of these categories. For instance, both Facebook and Twitter allow users to geotag content, check-in to places and to leave messages on a business' page or feed. Hence, it is increasingly important to understand the different privacy implications of sharing data on LBSNs. In the following section, we provide a systematic view of the different types of attackers and attacks on users' privacy in LBSNs, as well as mitigation strategies to help limit or prevent negative consequences of such attacks.

8.3.2 Privacy Threats and Protection Mechanisms

LBSN providers collect, process and store multiple types of users' data. By mining users' IP addresses, browser metadata, GPS coordinates, health data, photos, videos and audio recordings, service providers are in a unique position to capitalize on such wealth of information. Although there are techniques that allow online users to hide their true IP address—by connecting to proxies, VPNs and anonymization overlay

[25] https://www.fitbit.com/, last retrieved Dec 16, 2016.

networks such as Tor[26]—and protect their anonymity when browsing or searching for content [32], they lose effectiveness when used for LBSNs because users of these networks want to associate their account with the location information they provide.

For LBSN users, it is often challenging to assess the risks involved when sharing location and other type of information with LBSNs. One of the reasons is that people usually lack the awareness about the possible negative consequences due to the leak of sensitive data [104].

Most LBSNs offer users means to manage their data on the platform, usually through permission settings that allow users to specify the conditions under which certain types of data can be used and revealed to others [17]. While it is true that permissions are an important instrument for users to manage their privacy, their scope is often limited to other users of the LBSNs, with the assumption that the service provider is trusted to store and manage all of the data it collects—which is usually explained in a privacy policy. In addition to permission settings, Tsai et al. [122] identified different tools that can help users protect their privacy, which include blacklisting social contacts that should not access any of the users' information, using a restricted sharing approach where content would be visible to only a subset of a user's contacts, establishing "geo-fences", which are regions where location data should not be attached or shared with the LBSN [106], and by using time-based rules, as time is also highly-indicative of the type of visited place [87].

To reconciliate the different usability and privacy aspects in LBSNs, Carbunar et al. [17] have defined a set of five requirements that a LBSNs has to satisfy in order to preserve users' privacy. First, the LBSN should protect users' data from unauthorized access. Second, the privacy protection tools should not affect the functionality of the LBSN for the users. Third, they should enable providers and registered entities to be able to extract aggregate statistics and information that are relevant to their business. Fourth, privacy tools should minimize additional investments that need to be made to support them, both for the LBSN providers as well as for the registered entities. Fifth, such tools should minimize the additional effort that registered entities need to devote in order to use the LBSN, hence preserving its usability. The authors also note that some of these requirements can be contradicting, and that it can be challenging to satisfy all of them simultaneously.

To better understand possible attacks on users' privacy in LBSNs, in the following we characterize the adversary and the underlying assumptions of different mitigation strategies.

8.3.2.1 Adversaries, Threats and Solutions

We define as an *adversary* in a LBSN an entity (user, group, organization) that wishes to either (1) obtain access to data or derive information about a user or entity

[26] https://www.torproject.org/, last retrieved Dec. 16, 2016.

to which it does not have access, (2) modify that data, or (3) impersonate other users or entities. There are usually two types of adversaries in such systems, i.e., *internal* and *external*.

- *Internal adversary*: An internal adversary is an entity that has an existing relationship with the service provider. Examples of possible internal adversaries include other curious users of the LBSN who may want to infer additional information about victims by exploiting or misusing some functionality of the LBSN. It also includes curious or malicious employees of the LBSN who may try to access users' data without authorization, and the LBSN itself.
- *External adversary*: An external adversary does not have a direct relationship with the service provider. For instance, an external adversary could be a curious or malicious outsider or group who wants to steal personal data about users of the LBSN by either attacking the provider directly, or by aggregating data about potential victims from other sources that may be related to the victims. Examples of external adversaries include cyber-criminals, stalkers and groups who wish to steal users' data, or disrupt and block the service from functioning properly.

In addition to characterizing an adversary as either internal or external, privacy protection mechanisms in LBSNs are usually developed to counter a specific adversarial model, which can be either (a) *semi-honest* or (b) *malicious* [43, 58]. In the semi-honest model—also known as honest-but-curious—the adversary is assumed to follow the specified protocol but may try to learn information from the different operations it performs on the data. On the contrary, a malicious adversary can deviate from the specified protocol in any possible way, in order to maximize its success in reducing the users' privacy. In some scenarios, there could be an entity that is fully trusted by the user to execute the protocols correctly and not to reveal any personal information to adversaries. In such a scenario, the adversarial model can be either semi-honest or malicious, with the assumption that the trusted third-party (TTP) does not collude with the adversary and does not reveal any information to it.

An adversary might have several goals when conducting an attack. Vicente et al. [129] define four distinct categories of location privacy threats for LBSN users:

1. *Location privacy*: A location privacy loss occurs when the exact location of a LBSN user is revealed, and this create a significant concern for the users if it can be linked to her identity [48, 56], as it allows adversaries to infer other sensitive information such as the user's home and work locations, interests, political affiliations, and health issues [129]. We discuss location privacy threats and protection techniques in Sect. 8.3.2.2.
2. *Absence privacy*: Similarly to location privacy, absence privacy allows adversaries to learn that a user is not at a certain place during a given period of time. Although the possible consequences of this privacy threat are less obvious, there have been multiple cases where knowing that a user was not at a given place has enabled burglars to successfully rob several residences, multiple times (as described in Sect. 8.3) [19, 27]. We elaborate on absence privacy in Sect. 8.3.2.3.

3. *Co-location privacy*: A co-location privacy loss occurs when adversaries are able to infer the co-presence of multiple users at the same location at a given time. The type of privacy threat is of particular concern to users who either do not wish to reveal their presence at some location they consider to be sensitive, or who do not wish to reveal their proximity to other users. This type of threat is exacerbated due to the fact that, if the privacy settings are not properly configured, users can share their location and tag other users who might be unaware of this until it is too late. Co-location privacy can be extended to include the more generic aspect of interdependent privacy [8, 59, 95], in which the privacy of one user is threatened by the actions of other users on the LBSN. For brevity, we refer the reader to the individual articles for more information about interdependent privacy. We discuss co-location privacy threats in Sect. 8.3.2.4.
4. *Identity privacy*: An identity privacy loss occurs when it is possible for an adversary to link an account on a LBSN to a particular identity. Such threat is significant in many scenarios in which users wish to preserve their anonymity or pseudonymity with respect to other users and external parties. The loss of anonymity on such services can have devastating consequences: In the 2015, a data breach on a popular online dating site affected the account details of 35 million members, which resulted in hundreds of sentimental relationship being broken [81]. We discuss identity privacy in Sect. 8.3.2.5.

In addition to these four categories of threats, a fifth category started to become increasingly important thanks to the large growth of the number of devices capable of capturing fitness and activity data [61]. We therefore include the *activity privacy* category as well:

5. *Activity privacy*: An activity privacy loss occurs when an adversary is able to infer the type of activity that a user is doing at a given time. By using large-scale social media data, researchers have been able to accurately model the urban activities of individuals and to predict the sequence of activities only by relying on check-ins and geotagged posts on social media [51–53, 73, 77]. We discuss activity privacy threats in Sect. 8.3.2.6.

In order to tackle the privacy requirements and challenges in LBSNs, the research community has focused on several approaches based on different underlying techniques [63, 129]. One major category of privacy-preserving techniques are based on statistical methods that modify the reported location information in the space and time domains [99, 121]. In that category, we can find the following methods: (1) Query enlargement techniques [63], where instead of reporting the exact location of the user to the LBSN provider, the reported location is expanded to cover a larger geographic region, (2) fake or dummy location reports [121], where the users would report a set of fake locations together with the actual location in order to hide it among the dummy ones, and (3) progressive retrieval techniques that enable users to retrieve information by iteratively querying the provider [128]. In addition to statistical methods, another set of techniques rely on strong cryptography in order to design protocols to ensure that only the intended parties are able to obtain

Table 8.2 Categorization of different research works according to the adversary type (In: internal, E: external), adversarial model (M: malicious, S: semi-honest, T: trusted third-party), the goal of the adversary (L: location, Ab: Absence, C: co-location, Id: identity, Ac: activity) and the proposed or suggested privacy protection mechanism (spatial/temporal cloaking, elimination, fake data, cryptography)

Property	[74]	[49]	[64]	[143]	[105]	[40]	[75]	[70]	[57]	[102]	[103]	[76]
Adversary type	E	E	E	In/E	In	In/E	In/E	In	In	In/E	E	E
Adversary model	M	M	M	S	S	T	S	S	S	M	S	M
Goal	L	L	L	L	Ab	Ab	C	C	C	Id	Id	Ac
Privacy protection	Sp. cl.	Elim.	N/A	N/A	N/A	Sp&Tp. cl.	Crypt.	Crypt.	Crypt.	N/A	N/A	Sp. cl.

the information they require through secure computations, and nobody else [42]. For example, Private Information Retrieval techniques enable users to retrieve information without revealing what they are looking for to the provider [13].

In the following, we describe some examples of different privacy threats and proposed solutions for each of five different threat categories. Table 8.2 provides a summary of each of the works we present according to the different adversarial models, privacy threats and solution methodologies described (if any).

8.3.2.2 Location

Hereafter we discuss several techniques that threaten location privacy of LBSN users, which rely on one or several of the following data sources: users' location trajectories, textual content, location check-ins, social relationships and photo subjects.

With regards to location privacy, Li et al. [74] have recently conducted a study to measure the similarity between the real mobility pattern and the disclosed locations (through check-ins, for example) of LBSN users. Their results, based on a sample of 30 volunteers who have been providing their actual location samples as well as their disclosed locations on LBSNs, show that there is a substantial gap between the mobility pattern that can be extracted from the disclosed locations and the real mobility: Only in 16–33% of the cases, the authors were able to successfully derive the top-2 POIs (such as home and work). One possible reason is that it has been observed how users tend not to check-in at places considered to be "boring", such as home or work locations [142]. One of the implications of such a result is that even an adversary who has access to the location check-ins of a user on the LBSN—but not to the actual location trace—will have troubles in identifying the most visited POIs of a user with accuracy.

In contrast to studying location traces, Han et al. [49] focus on the textual analysis of Twitter posts in order to extract linguistic cues that can be linked to a specific location. In particular, the authors study the text and user profile information in

order to predict the city where the user is located. The intuition is that, for example, users in London should be more likely to tweet about *piccadilly* and *tube* than users in New York or Bejing. Hence, the authors focus on identifying a small set of location indicative words (LIW) in order to increase the geolocalisation accuracy of their machine learning algorithm, both on the regional as well as global scales. The analysis, conducted using multi-lingual tweets, shows that it is possible to correctly predict the city for 49% of English users, with a median error distance of just 9 km. To preserve privacy, the authors suggest that users should reduce the usage of LIWs, particularly gazetted terms, and to delete location-sensitive data from their profiles (such as location and time-zone information).

In addition to the text of a post, shared media can also provide useful information for adversaries who want to infer the location of LBSN users. For instance, Zheng et al. [143] use the real scene captured in a photo in order to infer whether it represents a home or vacation location. Their algorithm, based on a convolutional neural network, examines both the scene of the pictures and their geotags in order to infer a user's home location within a cell of 100×100 m. The algorithm is able to correctly predict the home location of a user with an accuracy of 71%, within a 70.7 m error distance. With the shrinking cost of computational resources and the availability of machine learning models accessible on the cloud,[27] it is becoming increasingly affordable to process not just metadata but also the content of media in order to improve the predictive performance of location-inference algorithms.

LBSNs usually allow users to establish relationships with others, either in a symmetric (friendship) or asymmetric fashion (follower-followee). Such social networks can also be used in order to infer the location of the users, even if they do not reveal their location information. By studying the social relationships on Twitter, Jurgens [64] develops an algorithm based on spatial label propagation that is able to infer the locations of a group of Twitter users who generate 74% of all the daily message volumes. The inference algorithm starts from a small number of known locations from which it assigns the most likely location to users whose location is unknown. The algorithm is able to correctly predict the location of 50% of the users in a Twitter-based social network within 10 km. Moreover, the same technique is also able to infer the locations of 50% of users on a different social network (Foursquare), within a 25 km error.

8.3.2.3 Absence

As opposed to location inference, the goal of an absence inference attack is to infer whether a user was not physically present at a place during a given period of time. This attack can also have serious consequences in a scenario where the absence from a place is considered as sensitive information. For example, the absence of

[27] https://cloud.google.com/prediction/, \https://aws.amazon.com/machine-learning/, last retrieved Dec. 17, 2016.

an employee from her workplace during work hours could lead to disciplinary measures from the employer. In contrast to location privacy, where the privacy loss occurs if a user can be located at a given point in time, Saini and El Saddik [105] argue that for absence privacy, it is more appropriate to model the privacy loss during a period of time, because the absence from one place at a given point in time does not necessarily imply that the user was not there or in the vicinity at a different but very close time instant. In a first attempt to formalize absence privacy, Freni et al. [40] proposed a set of definitions and techniques to preserve absence privacy, which rely on spatio-temporal generalization of the reported location. Such techniques rely on a trusted third-party, which is responsible for enforcing users' privacy preferences through the notion of an absence privacy region, where an adversary cannot exclude any point as a possible location of the user. By means of temporal delays when publishing geotagged information, the authors show how the effect on the quality of service is relatively modest (16–26 min of delay), and that it largely depends on the amount of other users currently in a given area, as well as their posting frequency.

8.3.2.4 Co-location

One popular example of co-location privacy threat is represented by services that offer to notify users when they are in physical proximity to other users of a LBSN, usually referred to as "nearby friend alert" [75]. Solutions to such challenge are mostly based on cryptographic primitives [29, 57, 70, 85, 91, 93, 145], relying on secure multi-party computation, cryptographic hashing or either public- or symmetric-key encryption. For instance, Li et al. [78] propose a protocol for nearby friend alert that allows users to trade accuracy with communication overhead. Based on the grid-and-hashing approach [110]—which partitions the space is grids and compares the signatures of such grids between users in order to discover if they are in the same grid—the authors design a flexible algorithm that finds an optimal placement of such grids that reduces by more than 50% the number of required grids as compared to a random placement, hence saving communication and computation costs for the users and the service provider. Mascetti et al. [85] propose two cryptographic protocols (Hide&Crypt and Hide&Hash), based on set-inclusion, that rely on location obfuscation and encryption in order to provide secure proximity detection functionality that preserves the location privacy of the users with respect to other users and the service provider. Kotzanikolauo et al. [70] improve upon existing protocols based on private-equality testing by designing a lightweight solution that can be run on resource-constrained mobile devices. Similarly, Hu et al. [57] propose a novel scheme relying on homomorphic encryption—a set of cryptographic techniques allowing computations on encrypted data—and geo-hashing to enable users to determine whether they are in proximity without revealing their location to other users or the service provider, and to perform spatial cloaking over encrypted geographic coordinates.

8.3.2.5 Identity

In order to identify users on LBSNs, Rossi and Musolesi [102] developed a set of techniques based on the study of location check-in data. By using a Bayesian probabilistic model that relies on the sequence of check-ins, the frequency of their occurrence and the social ties of the users, the authors are able to correctly infer more than 90% of the identities of online users on different datasets of check-ins from existing LBSNs. Unsurprisingly, the authors show how the more unique a GPS position is, the more effective is their algorithm in identifying the user with a small number of check-ins. In a follow-up work, Rossi et al. [103] characterize the types of venues that an adversary should monitor in order to maximize its success. The results, based on a large dataset of more than 1 million check-ins from 17 urban regions of the U.S., show that unsurprisingly the type of venues in the category "Residence" have the highest re-identification potential. However, more surprisingly, the authors discover that users with a high location entropy—which means that they visit more distinct types of venues more frequently than other users with lower entropy—are not necessarily the hardest to re-identify. The authors claim that this result indicates how it is the collective behavior of many users that influences the complexity of re-identification, rather than the individual user's behavior.

In addition to re-identification, demographics inference can also pose a threat to users' privacy. In the work by Li et al. [74], the authors show that demographics inference is quite successful as it exploits similarities between check-in traces of different users, despite a relatively poor performance in predicting actual location traces from check-in data. Specifically, their algorithm is able to infer features such as age, occupation, living place, gender and education level with an accuracy of 69.2%, 53.8%, 54.5%, 73% and 76%, respectively, on a sample of over 22,000 volunteers.

8.3.2.6 Activity

An adversary might be able to infer the activity of a LBSN user from the type of place (such as "restaurant") that corresponds to the reported location, or from the sequence of location reports, which can happen at different time and space granularities. Lian and Xie [76] design and evaluate a method to infer the activity, i.e., the type of place a user is at, based on GPS readings, time, user identification and other contextual information. In such scenarios, one main challenge is the scarcity of sufficient samples that can be used for the inference. In order to overcome this, the authors propose to use data from other users' check-ins, provided that their check-in histories are similar. By leveraging clustering and matrix factorization techniques, the authors show that by training on all users' check-in data, instead of training only on the victim's check-in data, the prediction performance is reduced by only up to 10% (weighted F1-score).

In order to reduce the possible search space for the different types of places a user is likely to visit next, Ye et al. [138] build a technique which uses a mixed hidden Markov model for a 2-step prediction. First, the model predicts the category of the place a user is likely to visit next, and then it predicts a location given the category. Their approach reduces the number of possible location candidates by a factor of 5.45 and improves location prediction accuracy by over 13% on a dataset extracted from the LBSN Gowalla (which is no longer active [16]). More recently, Yang et al. [137] proposed a fusion model that combines two separate inference methods, one for spatial data and one for temporal data. Similarly to [76], the authors rely on affinities between different users' temporal activities and the specificity of one activity at any given location, and they show how their solution achieves consistently good performance on three different datasets from two LBSNs (Gowalla and Fourquare), improving upon various baseline methods.

8.3.3 Open Research Challenges

In this section, we have described the different privacy aspects that are relevant in a LBSN. From threat formalization, adversarial models, privacy requirements and protection techniques, the research community has studied a wide array of problems that have yet to find a unified framework and solution. The availability of large amounts of digital data that we leave by interacting with online services, known as "digital footprints" [80], coupled with the shrinking cost of computation and cloud-based machine-learning solutions, are already enabling powerful inferences about people's lives and affections. In Sect. 8.3.2.2 we have described how deep neural networks are able to enhance the performance of location inference by processing images collected from a LBSN, which is nowadays feasible for every adversary with a minimal cost. With more and more machine learning models available, it is important to assess the amount of private user information that is leaked from the model parameters themselves [108]. Hopefully, novel protection mechanisms are being developed to provide provable privacy guarantees against such adversaries, by combining data separation and adding statistical noise during the training process [1]. The utility implications of such methods have yet to be fully assessed, but it is clear that the more data about users' location-related activities are available, the greater is the risk of a potential misuse of such data.

A related open challenge remains the definitive measurement of privacy loss in LBSNs [107]. Currently, there are multiple ways of measuring privacy [7, 24, 54, 132], and researchers have yet to find a unified framework for measuring it. Progress has been achieved in the area of location privacy, where a unified framework based on accuracy, correctness and uncertainty has been proposed and validated [109]. However, more research is needed in other dimensions of privacy in LBSNs, such as co-location, absence, identity and activity.

No matter how effective privacy protection mechanisms can be, they would not achieve their fullest potential unless they are delivering a coherent, simple and

functional user experience [21, 41, 67]. Managing privacy on LBSNs is nowadays challenging for many users, and the controls that are offered are often insufficient or too complex for most users to manipulate [41]. To help users feel more comfortable when sharing personal information on LBSN, better ways of presenting benefits and controls will have to be studied and developed [122], as well as clearer privacy policies that users can read and understand [10, 89].

8.4 Conclusion

Geospatial applications have witnessed a great revolution thanks to the development of modern collaborative technologies that enable users to both consume and contribute geographic information to the online community. In the first part of this chapter, we have introduced and discussed privacy issues that arise when location data is attached to different types of content shared with online services. We have introduced geo-spatial applications, such as interactive thematic maps, which can have a significant positive outcome for the people in scenarios including disaster relief efforts, transportation and urban resource management. However, we have also pointed out how geo-spatial information that is publicly accessible can also represent a source of privacy concern for citizens who might not want to have their locations associated with data that could be used in order to discriminate them or the places in which they live. In particular, we have shown how crime maps could be perceived both positively, when they increase transparency and awareness, as well as negatively, when they could influence the perception of property values in certain areas.

In the second part of this chapter, we have focused on a subcategory of geo-spatial applications, namely location-based social networks (LBSNs), as a recent phenomenon that has gained tremendous popularity among mobile users. To better understand the complex interaction patterns in such services, which comprise users, organizations and service providers, we have outlined a framework that enables researchers and practitioners to adopt a principled approach towards privacy threats and solutions. Such framework encompasses the network architecture, the threat categories as well as solution approaches. Although these categories cover several known attack goals, our analysis is not limited to the currently available solutions, as there are still important open questions that need to be addressed. We identified three research challenges that will benefit from a broader and systematic analysis in order to yield benefits for the users of LBSNs: big data processing with privacy guarantees, comparable and unified metrics across different privacy scenarios, and improved user experience through better and easier controls for managing privacy settings, as well as clearer notices related to the use and collection of users' data on LBSNs.

References

1. M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM, 2016.
2. A. Acquisti, I. Adjerid, and L. Brandimarte. Gone in 15 seconds: The limits of privacy transparency and control. *IEEE Security & Privacy*, 4(11):72–74, 2013.
3. A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International workshop on privacy enhancing technologies*, pages 36–58. Springer, 2006.
4. G. J. Annas. Hipaa regulations-a new era of medical-record privacy? *New England Journal of Medicine*, 348(15):1486–1490, 2003.
5. S. B. Barnes. A privacy paradox: Social networking in the united states. *First Monday*, 11(9), 2006.
6. V. Benson, G. Saridakis, and H. Tennakoon. Information disclosure of social media users: does control over personal information, user awareness and security notices matter? *Information Technology & People*, 28(3):426–441, 2015.
7. M. Bezzi. An information theoretic approach for privacy metrics. *Transactions on Data Privacy*, 3(3):199–215, 2010.
8. G. Biczók and P. H. Chia. Interdependent privacy: Let me share your data. In *International Conference on Financial Cryptography and Data Security*, pages 338–353. Springer, 2013.
9. I. Bilogrevic, M. Jadliwala, I. Lám, I. Aad, P. Ginzboorg, V. Niemi, L. Bindschaedler, and J.-P. Hubaux. Big brother knows your friends: on privacy of social communities in pervasive networks. In *International Conference on Pervasive Computing*, pages 370–387. Springer Berlin Heidelberg, 2012.
10. I. Bilogrevic and M. Ortlieb. If you put all the pieces together...: Attitudes towards data combination and sharing across services and companies. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5215–5227. ACM, 2016.
11. A. J. Blatt. The benefits and risks of volunteered geographic information. *Journal of Map & Geography Libraries*, 11(1):99–104, 2015.
12. A. J. Blatt. Data privacy and ethical uses of volunteered geographic information. In *Health, Science, and Place*, pages 49–59. Springer, 2015.
13. D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III. Public key encryption that allows pir queries. In *Annual International Cryptology Conference*, pages 50–67. Springer, 2007.
14. J. A. Bovenberg, B. M. Knoppers, A. Hansell, and K. de Hoogh. Exposing participants? population biobanks go geo. *European Journal of Human Genetics*, 2015.
15. T. Burghardt, E. Buchmann, J. Müller, and K. Böhm. Understanding user preferences and awareness: Privacy mechanisms in location-based services. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 304–321. Springer, 2009.
16. J. Cabalona. Gowalla Is Officially Shut Down.
17. B. Carbunar, M. Rahman, and N. Pissinou. A survey of privacy vulnerabilities and defenses in geosocial networks. *IEEE Communications Magazine*, 51(11):114–119, 2013.
18. S. Chainey and L. Tompson. Engagement, empowerment and transparency: publishing crime statistics using online crime mapping. *Policing*, 6(3):228–239, 2012.
19. C. Chan. Robbers Checked Facebook Status Updates To See When People Weren't Home.
20. Y. Chen and H. Kobayashi. Signal strength based indoor geolocation. In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 1, pages 436–439. IEEE, 2002.
21. C. D. Cottrill et al. Location privacy preferences: A survey-based analysis of consumer awareness, trade-off and decision-making. *Transportation Research Part C: Emerging Technologies*, 56:132–148, 2015.

22. J. W. Crampton. Cartographic rationality and the politics of geosurveillance and security. *Cartography and Geographic Information Science*, 30(2):135–148, 2003.
23. J. W. Crampton. *Mapping: A critical introduction to cartography and GIS*, volume 11. John Wiley & Sons, 2011.
24. R. Dayarathna. Taxonomy for information privacy metrics. *J. Int'l Com. L. & Tech.*, 6:194, 2011.
25. C. Dempsey. Privacy in gis issues, 2008.
26. S. N. Desk. How many CCTV Cameras are there globally?, June 2015.
27. B. Dickinson. Infographic: 80% of robbers check Twitter, Facebook, Google Street View.
28. T. Dienlin and S. Trepte. Is the privacy paradox a relic of the past? an in-depth analysis of privacy attitudes and privacy behaviors. *European Journal of Social Psychology*, 45(3):285–297, 2015.
29. W. Dong, V. Dave, L. Qiu, and Y. Zhang. Secure friend discovery in mobile social networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 1647–1655. IEEE, 2011.
30. S. O. Dyke, E. S. Dove, and B. M. Knoppers. Sharing health-related data: a privacy test? *NPJ Genomic Medicine*, 1:16024, 2016.
31. N. Eagle and A. S. Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.
32. E. Erdin, C. Zachor, and M. H. Gunes. How to find hidden users: a survey of attacks on anonymity networks. *IEEE Communications Surveys & Tutorials*, 17(4):2296–2316, 2015.
33. Facebook. Help Center - Location Privacy.
34. Facebook. Privacy Settings and Tools.
35. Facebook. Facebook Reports Second Quarter 2016 Results, 2016.
36. Facebook. Stats, facebook newsroom, 2016.
37. Fitbit. What should i know about my fitbit badges?, 2016.
38. M. Foucault. *Discipline and punish: The birth of the prison*. Vintage, 1977.
39. M. Foucault, G. Burchell, C. Gordon, and P. Miller. *The Foucault effect: Studies in governmentality*. University of Chicago Press, 1991.
40. D. Freni, C. Ruiz Vicente, S. Mascetti, C. Bettini, and C. S. Jensen. Preserving location and absence privacy in geo-social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 309–318. ACM, 2010.
41. M. Furini. Users behavior in location-aware services: Digital natives versus digital immigrants. *Advances in Human-Computer Interaction*, 2014, 2014.
42. G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan. Private queries in location based services: anonymizers are not necessary. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 121–132. ACM, 2008.
43. O. Goldreich. *Foundations of cryptography: volume 2, basic applications*. Cambridge university press, 2009.
44. P. Golle and K. Partridge. On the anonymity of home/work location pairs. In *International Conference on Pervasive Computing*, pages 390–397. Springer, 2009.
45. M. F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.
46. Google. Google My Account.
47. Google. The Next Chapter for Flu Trends.
48. M. Gruteser and B. Hoh. On the anonymity of periodic location samples. In *International Conference on Security in Pervasive Computing*, pages 179–192. Springer, 2005.
49. B. Han, P. Cook, and T. Baldwin. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*, 49:451–500, 2014.
50. F. Harvey. To volunteer or to contribute locational information? towards truth in labeling for crowdsourced geographic information. In *Crowdsourcing Geographic Knowledge*, pages 31–42. Springer, 2013.
51. S. Hasan and S. V. Ukkusuri. Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies*, 44:363–381, 2014.

52. S. Hasan and S. V. Ukkusuri. Location contexts of user check-ins to model urban geo life-style patterns. *PloS one*, 10(5):e0124819, 2015.
53. S. Hasan, S. V. Ukkusuri, and X. Zhan. Understanding social influence in activity-location choice and life-style patterns using geo-location data from social media. *Frontiers in ICT*, 3:10, 2016.
54. D. S. Herrmann. *Complete guide to security and privacy metrics: measuring regulatory compliance, operational resilience, and ROI*. CRC Press, 2007.
55. H. Hodson. Baidu uses millions of users' location data to make predictions, 2016.
56. B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5(4):38–46, 2006.
57. P. Hu, T. Mukherjee, A. Valliappan, and S. Radziszowski. Homomorphic proximity computation in geosocial networks. In *Computer Communications Workshops (INFOCOM WKSHPS), 2016 IEEE Conference on*, pages 616–621. IEEE, 2016.
58. Y. Huang, J. Katz, and D. Evans. Quid-pro-quo-tocols: Strengthening semi-honest protocols with dual execution. In *2012 IEEE Symposium on Security and Privacy*, pages 272–284. IEEE, 2012.
59. M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti. On non-cooperative genomic privacy. In *International Conference on Financial Cryptography and Data Security*, pages 407–426. Springer, 2015.
60. ICO. Crime-mapping and geo-spatial crime data: Privacy and transparency principles., 2012.
61. IDC. Worldwide Wearables Market Increases 67.2% Amid Seasonal Retrenchment, According to IDC.
62. N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, Aug. 2016.
63. C. S. Jensen, H. Lu, and M. L. Yiu. Location privacy techniques in client-server architectures. In *Privacy in location-based applications*, pages 31–58. Springer, 2009.
64. D. Jurgens. That's what friends are for: Inferring location in online social media platforms based on social relationships. *ICWSM*, 13:273–282, 2013.
65. M. Kandias, L. Mitrou, V. Stavrou, and D. Gritzalis. Which side are you on? a new panopticon vs. privacy. In *Security and Cryptography (SECRYPT), 2013 International Conference on*, pages 1–13. IEEE, 2013.
66. B. Kar and R. Ghose. Is my information private? geo-privacy in the world of social media. In *GIO@ GIScience*, pages 28–31, 2014.
67. H.-S. Kim. What drives you to check in on facebook? motivations, privacy concerns, and mobile phone involvement for location-based information sharing. *Computers in Human Behavior*, 54:397–406, 2016.
68. R. Kitchin. The real-time city? big data and smart urbanism. *GeoJournal*, 79(1):1–14, 2014.
69. R. Kitchin. Data-driven, networked urbanism. In *Data and the City workshop*, 2015.
70. P. Kotzanikolaou, C. Patsakis, E. Magkos, and M. Korakakis. Lightweight private proximity testing for geospatial social networks. *Computer Communications*, 73:263–270, 2016.
71. O. Kounadi, K. Bowers, and M. Leitner. Crime Mapping On-line: Public Perception of Privacy Issues. *European Journal on Criminal Policy and Research*, 21(1):167–190, Mar. 2015.
72. D. Lazer, R. Kennedy, G. King, and A. Vespignani. The parable of google flu: traps in big data analysis. *Science*, 343(6176):1203–1205, 2014.
73. J. H. Lee, S. Gao, K. Janowicz, and K. G. Goulias. Can twitter data be used to validate travel demand models? In *IATBR 2015-WINDSOR*, 2015.
74. H. Li, H. Zhu, S. Du, X. Liang, and X. Shen. Privacy leakage of location sharing in mobile social networks: Attacks and defense. *IEEE Transactions on Dependable and Secure Computing*, (1):1–1, 2016.
75. H. P. Li, H. Hu, and J. Xu. Nearby friend alert: Location anonymity in mobile geosocial networks. *IEEE Pervasive Computing*, 12(4):62–70, 2013.

76. D. Lian and X. Xie. Collaborative activity recognition via check-in history. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pages 45–48. ACM, 2011.
77. F. Liu, D. Janssens, J. Cui, G. Wets, and M. Cools. Characterizing activity sequences using profile hidden markov models. *Expert Systems with Applications*, 42(13):5705–5722, 2015.
78. Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 61–70. ACM, 2011.
79. M. Madden. Americans Consider Certain Kinds of Data to be More Sensitive than Others, Nov. 2014.
80. M. Madden, S. Fox, A. Smith, and J. Vitak. *Digital Footprints: Online identity management and search in the age of transparency*. Pew Internet & American Life Project Washington, DC, 2007.
81. B. M. C. f. MailOnline. Ashley Madison members reveal the impact of last year's hack, Aug. 2016.
82. A. Maiti, M. Jadliwala, J. He, and I. Bilogrevic. (smart) watch your taps: side-channel keystroke inference attacks using smartwatches. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 27–30. ACM, 2015.
83. S. Mardenfeld, D. Boston, S. J. Pan, Q. Jones, A. Iamntichi, and C. Borcea. Gdc: Group discovery using co-location traces. In *Social computing (SocialCom), 2010 IEEE second international conference on*, pages 641–648. IEEE, 2010.
84. M. Maron. How complete is OpenStreetMap?
85. S. Mascetti, D. Freni, C. Bettini, X. S. Wang, and S. Jajodia. Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies. *The International Journal on Very Large Data Bases*, 20(4):541–566, 2011.
86. D. McCullagh. Microsoft's Web map exposes phone, PC locations, 2011.
87. G. McKenzie and K. Janowicz. Where is also about time: A location-distortion model to improve reverse geocoding using behavior-driven temporal semantic signatures. *Computers, Environment and Urban Systems*, 54:1–13, 2015.
88. P. Meier. Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time. *Journal of Map & Geography Libraries*, 8(2):89–100, 2012.
89. G. R. Milne and M. J. Culnan. Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3):15–29, 2004.
90. MIT. Real Time Rome.
91. A. Narayanan, N. Thiagarajan, M. Lakhani, M. Hamburg, and D. Boneh. Location privacy via private proximity testing. In *NDSS*, 2011.
92. B. News. Ice age star map discovered, 2000.
93. J. D. Nielsen, J. I. Pagter, and M. B. Stausholm. Location privacy via actively secure private proximity testing. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 381–386. IEEE, 2012.
94. A. Noulas, M. Musolesi, M. Pontil, and C. Mascolo. Inferring interests from mobility and social interactions. In *NIPS Workshop on Analyzing Networks and Learning with Graphs*, pages 2–88, 2009.
95. A.-M. Olteanu, K. Huguenin, R. Shokri, M. Humbert, and J.-P. Hubaux. Quantifying interdependent privacy risks with location data. *IEEE Transactions on Mobile Computing*, 2016.
96. E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang. Accessory: password inference using accelerometers on smartphones. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, page 9. ACM, 2012.
97. R. L. Penslar and J. Porter. Institutional review board guidebook. *Retrieved July*, 18:2003, 1993.

98. Pew Research Center. More Americans using smartphones for getting directions, streaming TV, Jan. 2016.
99. N. Polatidis, C. K. Georgiadis, E. Pimenidis, and E. Stiakakis. A method for privacy-preserving context-aware mobile recommendations. In *International Conference on e-Democracy*, pages 62–74. Springer, 2015.
100. K. Pomfret. Latitudes and attitudes: Zooming in on geospatial data, privacy and the law in the digital age. *Centre for Spatial Law and Policy*, 2013.
101. J. H. Reiman. Driving to the panopticon: A philosophical exploration of the risks to privacy posed by the highway technology of the future. *Santa Clara Computer & High Tech. LJ*, 11:27, 1995.
102. L. Rossi and M. Musolesi. It's the way you check-in: identifying users in location-based social networks. In *Proceedings of the second ACM conference on Online social networks*, pages 215–226. ACM, 2014.
103. L. Rossi, M. Williams, C. Stich, and M. Musolesi. Privacy and the city: User identification and location semantics in location-based social networks. In *Ninth International AAAI Conference on Web and Social Media*, 2015.
104. Z. Sahnoune, C. Y. Yep, and E. Aïmeur. Privacy issues in geosocial networks. In *International Conference on Risks and Security of Internet and Systems*, pages 67–82. Springer, 2014.
105. M. Saini and A. El Saddik. Absence privacy loss. *Computer*, 48(11):102–105, 2015.
106. D. Schoepe and A. Sabelfeld. Understanding and enforcing opacity. In *2015 IEEE 28th Computer Security Foundations Symposium*, pages 539–553. IEEE, 2015.
107. K. G. Shin, X. Ju, Z. Chen, and X. Hu. Privacy protection for users of location-based services. *IEEE Wireless Communications*, 19(1):30–39, 2012.
108. R. Shokri, M. Stronati, and V. Shmatikov. Membership inference attacks against machine learning models. *arXiv preprint arXiv:1610.05820*, 2016.
109. R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux. Quantifying location privacy. In *2011 IEEE Symposium on Security and Privacy*, pages 247–262. IEEE, 2011.
110. L. Šikšnys, J. R. Thomsen, S. Šaltenis, M. L. Yiu, and O. Andersen. A location privacy aware friend locator. In *International Symposium on Spatial and Temporal Databases*, pages 405–410. Springer, 2009.
111. R. Soden and L. Palen. From crowdsourced mapping to community mapping: the post-earthquake work of openstreetmap haiti. In *COOP 2014-Proceedings of the 11th International Conference on the Design of Cooperative Systems, 27–30 May 2014, Nice (France)*, pages 311–326. Springer, 2014.
112. D. J. Solove. A taxonomy of privacy. *University of Pennsylvania law review*, pages 477–564, 2006.
113. H. Soroush, K. Sung, E. Learned-Miller, B. N. Levine, and M. Liberatore. Turning off gps is not enough: Cellular location leaks over the internet. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 103–122. Springer, 2013.
114. A. Staff. Location Sharing Is Easier Than You Think, 2016.
115. D. T. Staff. The history of social networking, May 2016.
116. B. Steele. NPR's Gravitas, Sept. 2002.
117. F. Stutzman, R. Capra, and J. Thompson. Factors mediating disclosure in social network sites. *Computers in Human Behavior*, 27(1):590–598, 2011.
118. F. Stutzman, R. Gross, and A. Acquisti. Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of privacy and confidentiality*, 4(2):2, 2013.
119. N. J. Thrower. *Maps and civilization: cartography in culture and society*. University of Chicago Press, 2008.
120. L. Tompson, S. Johnson, M. Ashby, C. Perkins, and P. Edwards. Uk open source crime data: accuracy and possibilities for research. *Cartography and Geographic Information Science*, 42(2):97–111, 2015.
121. M.-T. Tran, I. Echizen, and A.-D. Duong. Binomial-mix-based location anonymizer system with global dummy generation to preserve user location privacy in location-based services. In *Availability, Reliability, and Security, 2010. ARES'10 International Conference on*, pages 580–585. IEEE, 2010.

122. J. Y. Tsai, P. G. Kelley, L. F. Cranor, and N. Sadeh. Location-sharing technologies: Privacy risks and controls. *ISJLP*, 6:119, 2010.
123. U.S. Department of Commerce Economic and Statistics Administration, U.S. Census Bureau. Small area income and poverty estimates (saipe) program, 2014.
124. A. Uteck. Ubiquitous computing and spatial privacy. *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society*, pages 83–101, 2009.
125. S. Utz and N. Kramer. The privacy paradox on social network sites revisited: The role of individual characteristics and group norms. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 3(2):2, 2009.
126. VB Insight. Identity and marketing: Capturing, unifying, and using customer data to drive revenue growth | Insight | VentureBeat, 2015.
127. VCloudNews. Every Day Big Data Statistics – 2.5 Quintillion Bytes of Data Created Daily.
128. C. R. Vicente, I. Assent, and C. S. Jensen. Effective privacy-preserving online route planning. In *2011 IEEE 12th International Conference on Mobile Data Management*, volume 1, pages 119–128. IEEE, 2011.
129. C. R. Vicente, D. Freni, C. Bettini, and C. S. Jensen. Location-related privacy in geo-social networks. *IEEE Internet Computing*, 15(3):20–27, 2011.
130. B. Vladlena, G. Saridakis, H. Tennakoon, and J. N. Ezingeard. The role of security notices and online consumer behaviour: An empirical study of social networking users. *International Journal of Human-Computer Studies*, 80:36–44, 2015.
131. N. Vratonjic, K. Huguenin, V. Bindschaedler, and J.-P. Hubaux. How others compromise your location privacy: The case of shared public ips at hotspots. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 123–142. Springer, 2013.
132. I. Wagner and D. Eckhoff. Technical privacy metrics: A systematic survey. *arXiv preprint arXiv:1512.00327*, 2015.
133. M. Wall. Big Data: Are you ready for blast-off? *BBC News*, Mar. 2014.
134. S. D. Warren and L. D. Brandeis. The right to privacy. *Harvard law review*, pages 193–220, 1890.
135. J. Wartell and J. T. McEwen. Privacy in the information age: A guide for sharing crime maps and spatial data series: Research report. *Institute for Law and Justice, U.S. Department of Justice, Office of Justice Programs*, 2001.
136. D. Wyatt, T. Choudhury, J. Bilmes, and J. A. Kitts. Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):7, 2011.
137. D. Yang, D. Zhang, V. W. Zheng, and Z. Yu. Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1):129–142, 2015.
138. J. Ye, Z. Zhu, and H. Cheng. What's your next move: User activity prediction in location-based social networks. In *Proceedings of the SIAM International Conference on Data Mining. SIAM*. SIAM, 2013.
139. A. L. Young and A. Quan-Haase. Privacy protection strategies on facebook: The internet privacy paradox revisited. *Information, Communication & Society*, 16(4):479–500, 2013.
140. M. Yutaka, O. Naoaki, I. Kiyoshi, et al. Inferring long-term user property based on users' location history. In *Proceedings of the IJCAI*. IJCAI, 2007.
141. P. A. Zandbergen. Accuracy of iphone locations: A comparison of assisted gps, wifi and cellular positioning. *Transactions in GIS*, 13(s1):5–25, 2009.
142. Z. Zhang, L. Zhou, X. Zhao, G. Wang, Y. Su, M. Metzger, H. Zheng, and B. Y. Zhao. On the validity of geosocial mobility traces. In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks*, page 11. ACM, 2013.
143. D. Zheng, T. Hu, Q. You, H. Kautz, and J. Luo. Towards lifestyle understanding: Predicting home and vacation locations from user's online photo collections. In *Proceedings of the 9th International AAAI Conference on Web and Social Media*, pages 553–560, 2015.
144. Y. Zheng. Location-based social networks: Users. In *Computing with spatial trajectories*, pages 243–276. Springer, 2011.

145. G. Zhong, I. Goldberg, and U. Hengartner. Louis, lester and pierre: Three protocols for location privacy. In *International Workshop on Privacy Enhancing Technologies*, pages 62–76. Springer, 2007.
146. M. Zook, M. Graham, T. Shelton, and S. Gorman. Volunteered geographic information and crowdsourcing disaster relief: a case study of the haitian earthquake. *World Medical & Health Policy*, 2(2):7–33, 2010.