1-1-2015

Location Privacy in the Era of the Internet of Things and Big Data Analytics

Robert P. Minch
Boise State University

Location Privacy in the Era of the Internet of Things and Big Data Analytics

Robert P. Minch
Boise State University
rminch@boisestate.edu

Abstract

Location information is generated in large quantities in the Internet of Things and becomes a major component of the big data phenomenon. This results in privacy issues involving sensing, identification, storage, processing, sharing, and use of this information in technical, social, and legal contexts. These issues must be addressed if the IoT is to be widely adopted and accepted. Theory will need to be developed and tested, and new research questions will need to be investigated. This exploratory research begins to identify, classify, and describe these issues and questions.

1. Introduction

The interactions of The Internet of Things (IoT) and big data analytics are greatly impacted by location information and in turn greatly impact location privacy. We will examine the relationships between these four concepts with an aim toward furthering a framework for future analysis and research.

Location is a critical and often central component of context-aware computing [1] and similar notions of ubiquitous and pervasive computing [2]. In the past location privacy was of relatively little concern because location information was not pervasively and continuously available. Now that technology has radically altered information availability, privacy of location is closely tied to controlling access to this information, and people want to be in control of the information availability [3]. Location privacy preferences are now quite well studied in the context of users carrying mobile devices [4] but not extended through an IoT context where device-to-device communication can carry location information far beyond users' awareness.

Privacy concerns are becoming an increasingly critical issue in the IoT [5]. Without assurance of privacy in a world of interconnected sensors and systems, users will be unwilling to adopt these new technologies [6]. The International Telecommunications Union report on the Internet of Things notes that "Concerns about privacy and data protection are widespread, particularly as sensors and smart tags can track a user's movements, habits, and preferences on a perpetual basis." [7]

Despite its relevance and importance, privacy is not yet receiving adequate attention in the enthusiasm to exploit the technical capabilities of the IoT. A recent survey of IoT literature covering 127 journal and conference papers [8] finds only nine security and three privacy-related documents in its category of IoT challenges [9] [5] [10]. A recent survey of IoT context-aware computing describes security and privacy as a major concern, yet finds only 11 of 50 surveyed research prototypes incorporating security and privacy functionality [11].

The remainder of the paper is organized as follows. Sections 2 and 3 will review the IoT and big data as needed for following discussion. Section 4 will discuss theory and methods. Sections 5 and 6 will address location information and location privacy, respectively. Section 7 will discuss the implications of issues raised and propose questions for further research. Section 8 is a summary.

2. The Internet of Things

Said to be coined as a phrase in 2009 [12] the IoT recognizes that the fastest-growing mode of communication on the Internet is not between communicating people but rather communicating "things"—sensors, actuators, and other devices and objects. Early work on the IoT emphasized RFID as a source, and supply chain networks as the application, often defining IoT in similar terms, e.g., "The 'things' are physical objects carrying RFID tags with a unique EPC" [13] and "The Internet of Things (IoT) is an emerging global Internet-based information architecture facilitating the exchange of goods and services in global supply chain networks." [13] More recently its context and scope have been considerably expanded and the IoT's core concept is said to be that ". . . everyday objects can be equipped with identifying, sensing, networking, and processing capabilities that will allow them to communicate with one another and with other devices and services over the Internet to achieve some useful objective." [8]

The IoT is an evolving construct. The networking firm Cisco has proposed an even broader concept called the Internet of Everything (IoE) [14]. The IoE is said to connect people, data, processes, and things in a network that combines machine-to-machine, person-to-machine, and person-to-person communication. Cisco predicts US $19 trillion in economic value for the IoE by 2020. Applications are seen in retail, manufacturing, and the public sector. Also building upon a IoT base is the proposed Web of Things (WoT) [15]. The WoT extends the IoT by exploiting Web technologies and standards to bring together a wide variety of smart devices into "physical mashups," including lightweight ad-hoc systems that may be rapidly created and torn down as needed.

An example of recently introduced components in the IoT are devices aimed at home automation and control. In June 2014 Honeywell International introduced a Wi-Fi, Internet-accessible thermostat [16] controllable by smart phone and compatible with Apple's HomeKit [17]. Honeywell also makes smoke detectors, window sensors, and other connectable devices, and Apple's HomeKit is intended to promote interoperability between multiple vendors and numerous devices including lights, locks, and cameras. This comes partially in response to the innovative Nest thermostat [18], and suggests a vibrant and competitive market for not only devices but standards and protocols for interoperability.

3. Big data analytics and management

Big data is often described in terms of Vs—Volume, Velocity, and Variety [19]. Some add additional Vs such as Veracity and Value [20] [21]. Volume, Velocity, and Variety issues and examples are briefly reviewed below.

3.1. Volume

The worldwide volume of data creation is increasing dramatically, from an estimated 1 zettabyte (10^{21} bytes) in 2010, to 7 zettabytes in 2014, possibly to reach 35 zettabytes by 2020 [22]. Until now, much of this data was created through relatively conventional means—for example, Walmart creates more than 2.5 petabytes of data per hour from customer transactions [19]. New sources of data from "edge" devices such as sensors and smartphones are now contributing to the rapid increase in data sources.

Examples of large data volumes abound. In 2012 the company Renew installed sensors in London recycling bins which recorded MAC addresses of phones and other devices with Wi-Fi turned on [23]. In a trial of 12 such bins, on the single day of July 6, 2012, 946,016 presences were detected, corresponding to 106,629 unique MAC addresses and presumably a correspondingly large number of passersby (although an individual person may carry more than one device with a MAC address and there are issues involved in mapping devices to persons). [23] A company spokesman estimated that 80% of people leave Wi-Fi on when leaving their home or office. In a salient example of inferencing capabilities possible through multiple sensors, the company proposed placing sensors in rest rooms, so that each MAC address could be associated with the gender of the person carrying the device. In August 2013 the City of London requested that the trail be halted, although the company claimed the devices recorded only "extremely limited, encrypted, aggregated, and anonymized data."

Modern smart electricity meters can collect high-resolution energy consumption data, and this can facilitate identification and usage patterns of household appliances. Already this has been exploited to the extent that specific television programs have been detected by monitoring the electrical usage patterns caused by various display technologies that vary power consumption according to material shown. [24] One captured communication between this smart meter and a utility company server contains 955 characters (an HTTP 1.1 POST with Content-Length 851, plus header information, uncompressed) [24]. The data is sent every two seconds. Assuming eight bits per character, the resulting bandwidth is 955 * 8 / 2 or 3820 bps. Storing the data would result in approximately 40 MB per user per day, and 15 GB per year. If implemented on the more than 200 million households in the EU, annual storage would be approximately three exabytes ($3 * 10^{18}$ bytes). This can be compared to an estimated 2.5 exabytes of global data created each day in 2012, or about 900 exabytes per year. [19]

Of course recent revelations about data collection by the US National Security Agency have also brought to light the unprecedented volumes of data that can and are being collected by national governments. In 2012, there were at least 41 billion total records collected and stored by XKeyscore (a system that intercepts Internet traffic including emails, Web addresses, social media posts, and more) in a single 30-day period. [25] It is speculated that the total storage at one new million-square-foot NSA data center in Utah will hold one yottabyte (10^{24} bytes, equivalent to one thousand zettabytes or one million exabytes) or more in storage [26].

3.2. Velocity

Both people and devices are producing data at unprecedented rates today. While data rates (volume per time period) are important, the latency and rapid possible response times are equally critical. For example, the Twitter social networking service averages some 6000 tweets per second, with a recent record of over 143,000 tweets in a single second on August 3, 2013 [27]. This means that analysis of topical trends can updated on a sub-second response time.

Data rates from devices vary greatly depending on the type and precision of information in use—even relatively simple automotive GPS systems provide location updates several times per second. Movement tracking is an example of how a sequence of location fixes can result in high-velocity information streams as the changing locations can also lead to the updating of maps and other information in the local context.

High velocity also supports continuous, real-time processing and decision making in business. Historically credit card companies might batch together occasional offers to potential customers via mass mailings prepared over several weeks. Today offer processing time has been reduced to a day, and in cases where Web and call center activity can be monitored continuously, personalized offers can be made in milliseconds [28].

3.3. Variety

Variety of big data is evident starting with sources of data, including sensors, appliances, social network sites, mobile phones, and many other devices with sufficient connectivity to share and store data. While commercial examples abound, it is also instructive to consider examples that begin as simple consumer goods but expand based on the addition of an ever-enlarging set of components and interactions between them. One such system is IFTTT.

IFTTT ("If This Then That") [29] is a simple rule-based system that has become popular for controlling the interaction of smart and connected devices. Various "channels" have been defined for particular devices and services, each with antecedent "triggers" and consequent "actions." Simple programs called "recipes" relate the triggers and actions. For example, one recipe specifies that an incoming email from a specific address will cause a Philips Hue connected light to blink as a notification. To date there are already more than 100 IFTTT channels [30].

IFTTT recipes can be chained together, leading to potentially unpredictable control and information sharing scenarios. It is notable that even simple recipes can introduce risk if used incautiously—for example, there is a shared IFTTT recipe triggered by a Netatmo temperature sensor that will turn on a space heater plugged into a Belkin WeMo remote-control electrical outlet. [31]

It is also becoming more common to integrate physical devices and sensors with human social networks in ways where each supplies input to the other [32] [33]. GPS location information in mobile phones and automobiles is generated on a physical device, but is typically most relevant as a proxy for a person's location, and may automatically feed into human social networks. In the other direction posts to a social network such as a stated intent to leave work early may lead to actuation of physical devices such as adjusting a home thermostat. Existing systems such as Google Now [34] monitor a user's email, Facebook, location, airline reservations, and more in order to predict the user's immediate and future needs.

3.4. Analytics and management

There are many implications of data analytics for privacy, for location analysis, and for the combination of the two, affecting location privacy. The length of time that consumers are willing to travel to shopping malls, for example, has been used to measure consumer demand [35]. Interpersonal social relationships and friendships have been accurately predicted by monitoring the locations of mobile phones [36]. The ubiquity of the IoT has also resulted in a call for its integration with cloud computing, resulting in the Cloud of Things (CoT) [37] [38]. Quantities of devices and volumes of data are said to be too large and too valuable to only store locally, therefore the CoT will allow for global utilization. Sensing as a Service (SaaS) is proposed as facilitating a ubiquitous computing environment, further enhancing the volume and availability of IoT data [21]. Of course global availability implies possible global exposure, therefore adding to the importance of security and privacy safeguards [37].

4. Theory and methods

The present work is an informational component in the early stages of theory development and in the broader context of the scientific process [39] rather than application of an existing theory or hypothesis testing. It follows and borrows from prior work investigating issues in the development of location privacy theory in general [40], and extends it to the particular environment of the IoT.

Theory construction has five desirable goals (reproduced from [41]):

1. A method of organizing and categorizing "things," a typology;
2. Predictions of future events;
3. Explanations of past events;
4. A sense of understanding about what causes events; and occasionally mentioned as well:
5. The potential for control of events.

At this early stage we can hardly purport to fully explain and predict the eventual evolution of the recently-emerged IoT, let alone control it—however we can begin to organize and categorize important "things" such as components and concepts.

Consistent with the above, a number of strategies may be used to construct theories, one of which is a classificatory strategy seeking a taxonomy of elements both within and outside the phenomenon [39]. In early stages of theory construction, classification strategies are particularly important and a prerequisite to other strategies [42]. This approach follows recommendations from related fields [43], emphasizing discovery and description, where key research questions are "Is there something interesting enough to justify research?" and "What are the key issues?" in both cases with categorization suggested as a procedure to be used [43] (page 324).

The methods described below will attempt to discover, classify, and describe a number of key issues that relate the IoT and big data to location privacy, and justify the need for additional research.

5. Location information

While early and basic IoT approaches began with networks of sensors and actuators, for our purposes it is useful to flesh out the possible information flows and address them in somewhat more detail. System-wide information privacy cannot be ensured only through controls on individual elements, but rather must be considered in light of overall communication flows. Table 1 shows six phases of location information flow, along with example components and methods associated with these. The phases are only partially ordered; sensing necessarily precedes identification, but storage, processing or sharing can follow in any order, or direct use of the information may be made with little storage, processing, or sharing. The phases are briefly described below.

- Location sensing may be accomplished by four main methods. Triangulation is the method used by cell towers (in the absence of GPS) with angular and time-of-arrival information. Scene analysis uses cameras and pattern recognition. Proximity uses signals such as Bluetooth or Wi-Fi to sense nearby transceivers. Indirect inference deduces location by observing correlated activities and events such as a garage door opening implying presence of an automobile and driver.
- Identification may be accomplished by detection of various unique identifiers. These may be Bluetooth or Wi-Fi addresses (e.g., MAC addresses), or any of a large number of standards-based or proprietary identifiers such as an iPhone unique identifier (UDID), Android device ID, etc. Availability of this ID information via broadcast or query depends upon several factors including vendor implementation details and user settings and preferences. Of course each of these methods depends upon a consistent mapping of sensors and devices to persons in order to identify human individuals and groups. A more direct form of identification is facial recognition, which can scan large crowds and in some cases has reached higher performance levels than human counterparts [44]. The location of recognized individuals is imputed from the location and viewing characteristics of the camera(s) used.
- Stored location information is itself considered metadata in some contexts such as photographs, however it has its own associated metadata as well. This may include absolute or relative external references, estimated accuracy, and more. It is outside the scope of the present work to discuss all implications of the object/meta aspects of such information, but it is worthy of note that the relative importance of each is not fixed but rather dependent on context (for example—in an emergency telephone call, either the caller's description of her location or the reported GPS location of the phone may be more valuable, depending on the particular context).
- Processing of location information may be relatively self-contained (e.g., a GPS-enabled device computing its own location, direction, and speed from received satellite signals) or it may depend on communication with additional systems and data analytics. In cases of multiple and/or communicating systems, the possibilities are very large—as is evident from police reports of tracking fugitives via locations of credit card charge, convenience store security cameras, toll-booth license plate scanners, and much more.
- Sharing of location information may be intentional or unintentional. Intentional sharing may be accomplished through apps such as Apple's Find my iPhone and Android's Find my Friends apps. Unintentional sharing is undoubtedly much more common, as a large majority of mobile device

users are not able or concerned with ensuring that their privacy settings are configured to prevent unintended sharing of information.

- Use of location information may likewise be classified into broad categories of intentional or unintentional use. Customers of car rental agencies may be pleased with the capabilities of their GPS navigation systems for way finding, but upset if they receive a speeding ticket due to the system detecting their vehicle speed and comparing it to posted speed limits at their various locations.

5.1. Increasing significance of location information

Strong trends in the IoT include further enlargement of the type and number of devices, greater affordability and marketplace availability of devices, and incorporation of sensing capabilities in devices that are already sold in large numbers. Apple's iBeacon is a recently-announced capability that can be enabled in many existing Apple iPhones without additional hardware, and can locate users of compatible devices within a few inches, using Bluetooth low-energy [45]. Apps in iOS 7 or later can define regions based on proximity to beacons (generally but not always stationary devices) and regions associated with those apps are monitored continuously, even when the app is not running [46]. It is estimated that half the top 100 US retailers will begin testing iBeacon beacons in their stores in 2014, and that beacons will be in more than 30,000 indoor locations by the end of 2014. Beacons will cost as little as $2 each by 2015 [45].

It may not be widely appreciated that stationary devices often have associated location information, and that this information is potentially useful, possibly sensitive in a security and privacy sense, and potentially one link in an inferential chain that may have much wider privacy implications. For example, most computer networking devices such as switches and routers hold a standardized MIB (Management Information Base) variable called sysLocation, which is defined as "The physical location of this node, (e.g., 'telephone closet, 3rd floor') [47]. This variable is typically queried via SNMP over a network. While device locations are necessary and useful for finding and repairing equipment, the location information itself clearly should be kept private for security reasons.

Most wireless access points may be configured with this MIB location information, and hence the location of associated users of mobile wireless devices such as laptop computers may be inferred via the location of their associated wireless access point. Similarly, other fixed devices such as smart thermostats are not necessarily inherently location aware, however if they connect to a Wi-Fi network their location may be inferred from the location of the wireless access point. Furthermore the location of devices such as smartphones they interact with may be location-aware, and the relationships between the devices may be exploited in location-aware applications. For example, iOS 8 and HomeKit can take advantage of "geofencing" [48] to set a thermostat to "home" mode when a homeowner is within a certain perimeter of the home, and "away" when outside that perimeter. Thus we see a chain of many links, passing location information among devices, users, and applications.

6. Location privacy

Location privacy may be viewed from many conceptual perspectives [40] and in the context of the present work related to the IoT and big data, we will consider it from an informational privacy perspective. The informational perspective is key to most privacy theories in a technological context, describing privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." [49] (p. 7). In keeping with this approach, we will look at location privacy in terms of information flows, from sensing to use and including a number of other activities typically in between (including more complex interactions between flows).

Table 2 uses the six phases of information flow enumerated in Table 1 and identifies example privacy controls for each phase. The six phases extend early work from more than 45 years ago identifying three phases of input, storage and output [49]. They also extend five phases discussed in [40] by explicitly adding the "processing" phase to acknowledge the important of inferencing capabilities and data analytical techniques that may deduce location from other available evidence.

The privacy-enhancing controls fall into technical, social, and legal measures, represented in columns of the table. Technical controls are those that control the actual processing of the information and may block, filter, modify, etc. that information. Examples include authenticating, blocking, encrypting, and other privacy protections for RFID tags [50]. Social controls affect privacy information through the influence of accepted business practices, social norms, and similar non-technical means. These include not only such things as formal privacy policies from system providers, but also behaviors of system users, which have been found to vary considerably according to context such as who the

information is exchanged with, whether the person is at home or in a public place, and what means is used to share the information [51]. Legal measures are those that impose formal prohibitions or regulations on activities related to location information flows. These vary greatly by region. In the EU the ePrivacy Directive directly addresses location privacy, while in the US federal law addresses location only indirectly and incompletely [52].

The six information flow phases and associated privacy controls are described below. Selected references are included in the text and in table cells.

- Sensing may be technically blocked by any means that prevents signal transmission or reception. This includes RFID-blocking wallets, RF blocker tags generating simulated or false RFID tags, etc. Social norms may also inhibit sensing, merely through the opprobrium of those who object to the panopticon scenarios made possible through technologies such as Google Glass. A number of laws prohibit the use of sensing-capable devices such as cameras and cell phones at border customs stations, locker rooms, and other sensitive locations.
- Identification in the IoT has already received significant attention [10] [5]. A recent development of interest for technical identification privacy is the announcement that Apple's iOS 8 will identify itself with a software-generated, random MAC address rather than the more conventional approach of a unique and fixed address [53]. This will significantly inhibit tracking systems that rely on recognizing the same MAC address in multiple locations. Social norms still often permit and on occasion encourage anonymity in letters to newspapers and postings to online discussion forums. Legal enforcement of anonymity is almost universally expected and enforced in particular contexts such as election ballot casting.
- Storage privacy is enabled through several technical methods, including merely not providing a storage facility and encryption of any stored data. The Snapchat service was touted as an ephemeral means of photo sharing, but was quickly and easily defeated [54]. Social control of stored information is often accomplished (with varying degrees of success and user satisfaction) through user privacy settings in social media. Various jurisdictions may enforce formal legal restrictions on the type, amount, and duration of stored data. A "right to quantitative privacy" has even been proposed [55].
- Processing phase technical privacy includes a number of design principles that also apply to other phases [56] and various anonymizing and privacy-enhancing and privacy-preserving technologies [5]. It may also be affected on a social and free market level in software terms of service agreements. Formal legal measures include prohibitions or restrictions on database matching and sharing of information between commercial entities.
- Sharing phase privacy may be technically implemented by restricting the communications channels available, e.g., not implementing or turning off facilities such as Bluetooth and Wi-Fi. Social measures are largely the responsibility of users to control application settings and follow recommended norms for appropriate sharing. Legal controls for sharing have recently received significant attention—for example the US Federal Trade Commission has just recommended that Congress give consumers more control over the data brokerage industry [57] and European courts have required that search engines implement a "right to be forgotten." [58]
- Use of privacy-sensitive information can be technically controlled partly by blocking its importation into applications. Most modern mobile device operating systems allow the user of an app to permit or deny the use of geo-location information by the app. Social controls on sensitive information are often governed by acceptable business practices and local norms. Legal restrictions are sometimes codified to prohibit discrimination, e.g., to prevent the practice of "redlining" where banks discriminate in granting mortgage loans based on locations and neighborhoods of potential borrowers.

There are interactions and inconsistencies in the above areas. For example, UK law allows anonymous collection of MAC addresses, but restricts the use of Web cookies unless the user gives permission. This leads to inconsistent net effects, since MAC address tracking is, in the view of the company tracking users via London recycling bins, "a cookie for the real world." [59] With multiple sensors, device location tracking can reveal much more than a simple cookie, including the speed at which a user is traveling, how long they pause at a particular location, and detailed movement patterns.

7. Discussion

New technologies and their uses have always had complex economic, social, cultural, and legal implications, with accompanying concerns about

negative consequences. The newly-invented telephone was once studied in an attempt to determine whether it might make men lazy and break up home life [60]. Yet it was not banned, gained near universal adoption, created remarkable efficiencies and conveniences for business and the public, and did not result in societal chaos. So it will probably be with the Internet of Things, big data, and their use of location data and attendant location privacy concerns.

Location information is a major component in effective inventory and supply chains, in efficient transportation systems, in context-aware mobile applications, and numerous other systems. These advantages must be balanced against security concerns, loss of privacy, and potential abuse made possible by the technologies involved. In part it is a question of whether a change in quantity is a change in quality, and whether a change in degree is a change in nature. Some will argue that a traffic light camera that records the license plates of speeders (and hence their location and speed information as well) is merely the substitution of a camera for a police officer with a radar gun and ticket book. Others will judge it a quantum difference, with continuous mass data collection posing a threat to personal rights and freedoms.

If one is concerned about location privacy in the context of the IoT and big data, then it's appropriate to consider how the situation should be managed. Because of the partial ordering of phases we have discussed, where for example sensing must occur before identification, if one believes that technical means are a good approach to the problem, then one might favor the ability to configure devices in the IoT so that they cannot collect location information at all. If not collected, it cannot be shared; if not collected there is no further need to be regulated. This approach favors controls in the top left cell of Table 2.

If one believes that location information's benefits generally (even if not universally) outweigh their costs, or that the collection, storage, sharing, etc. of location information is inevitable, then one might favor moving to the opposite cover of Table 2 (bottom right) and concentrate efforts on legal controls of location information use. This would be somewhat analogous to prohibiting health insurers from considering pre-existing conditions in patients—the information is known to all parties but its use in particular contexts is strictly prohibited.

It must be recognized that management and control of location information privacy may not be sufficient according to traditional user and public preferences. Society may need to balance the benefits of increased capabilities and efficiencies of the IoT against a possibly inevitably increased visibility into everyday business processes and personal activities. Much as people have come to accept increased sharing of personal information on the Web in exchange for better shopping experiences and other advantages, they may be willing to accept increased prevalence and reduced privacy of location information.

7.1. Questions for further research

Many location privacy research questions suggest themselves as the IoT evolves, and many of them align well with the information flow phases and privacy enhancing controls discussed above. Example questions, each of which may be translatable into families of formal hypotheses include:

1. What are the levels of user *awareness* concerning location information associated the IoT, and what affects these levels?
2. What are the levels of user *concern* about disclosure and use of location information associated with them, and what affects these?
3. What are the types and levels of *behaviors* adopted by users as a result of their location privacy concerns, and why are these adopted?
4. What are the *differences* in awareness, concerns, and behaviors of the general public versus business entities (e.g., participants in supply chains) related to location privacy?
5. For particular privacy concerns in particular contexts, what are the *preferred controls* for location privacy?
6. For particular privacy concerns in particular contexts, in what *phase of information flows* is control preferable?
7. What location privacy issues most strongly affect users' *willingness to adopt* particular IoT devices and features?

8. Summary

The Internet of Things and big data represent an explosion of information creation, sharing, and use. This is due to greatly increased types and numbers of connected physical devices such as sensors and actuators, and systems such as social networks used by people. Because location information is a large component of IoT information, and concerns about its privacy are critical to widespread adoption and confidence, location privacy issues must be effectively addressed.

It is hoped that the framework presented here, which looks at six phases of location information flow in the IoT and three areas of privacy controls that may be considered to manage those flows, will be helpful to

practitioners and researchers when evaluating the issues involved as the technology advances.

9. References

[1] B. N. Schilit and M. M. Theimer, "Disseminating active map information to mobile hosts," *IEEE Netw.*, vol. 8, no. 5, pp. 22–32, Sep. 1994.

[2] A. Smailagic and D. Kogan, "Location sensing and privacy in a context-aware computing environment," *IEEE Wirel. Commun.*, vol. 9, no. 5, pp. 10–17, Oct. 2002.

[3] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *Pervasive Comput. IEEE*, vol. 2, no. 1, pp. 46–55, 2003.

[4] M. Benisch, P. G. Kelley, N. Sadeh, and L. F. Cranor, "Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs," *Pers. Ubiquitous Comput.*, vol. 15, no. 7, pp. 679–694, Oct. 2011.

[5] V. Oleshchuk, "Internet of things and privacy preserving technologies," in *1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology, 2009. Wireless VITAE 2009*, 2009, pp. 336–340.

[6] C. Mayer, "Security and Privacy Challenges in the Internet of Things," 2009. .

[7] "ITU Internet Reports 2005: The Internet of Things." [Online]. Available: http://www.itu.int/osg/spu/publications/internetofthings/. [Accessed: 15-Jun-2014].

[8] A. Whitmore, A. Agarwal, and L. D. Xu, "The Internet of Things—A survey of topics and trends," *Inf. Syst. Front.*, pp. 1–14, Mar. 2014.

[9] C. M. Medaglia and A. Serbanati, "An Overview of Privacy and Security Issues in the Internet of Things," in *The Internet of Things*, D. Giusto, A. Iera, G. Morabito, and L. Atzori, Eds. Springer New York, 2010, pp. 389–395.

[10] A. C. Sarma and J. Girão, "Identities in the Future Internet of Things," *Wirel. Pers. Commun.*, vol. 49, no. 3, pp. 353–363, May 2009.

[11] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context Aware Computing for The Internet of Things: A Survey," *IEEE Commun. Surv. Tutor.*, vol. 16, no. 1, pp. 414–454, First 2014.

[12] K. Ashton, "That 'internet of things' thing," *RFiD J.*, vol. 22, pp. 97–114, 2009.

[13] R. H. Weber, "Internet of Things – New security and privacy challenges," *Comput. Law Secur. Rev.*, vol. 26, no. 1, pp. 23–30, Jan. 2010.

[14] "Internet of Everything," *Cisco*. [Online]. Available: http://www.cisco.com/web/about/ac79/innov/IoE.html. [Accessed: 15-Jun-2014].

[15] D. Guinard, V. Trifa, F. Mattern, and E. Wilde, "From the Internet of Things to the Web of Things: Resource-oriented Architecture and Best Practices," in *Architecting the Internet of Things*, D. Uckelmann, M. Harrison, and F. Michahelles, Eds. Springer Berlin Heidelberg, 2011, pp. 97–129.

[16] T. Black, "Thermostat War Heats Up as Honeywell Takes Aim at Google," *Bloomberg*, 10-Jun-2014. [Online]. Available: http://www.bloomberg.com/news/2014-06-10/thermostat-war-heats-up-as-honeywell-aims-at-google.html. [Accessed: 11-Jun-2014].

[17] "Apple's HomeKit will bring smart home control to iOS 8," *Macworld*, 02-Jun-2014. [Online]. Available: http://www.macworld.com/article/2357527/apples-homekit-will-bring-smart-home-control-to-ios-8.html. [Accessed: 11-Jun-2014].

[18] "Life with Nest Thermostat," *Nest*. [Online]. Available: https://nest.com/thermostat/life-with-nest-thermostat/. [Accessed: 14-Jun-2014].

[19] A. McAfee, E. Brynjolfsson, and others undefined, "Big data: the management revolution," *Harv. Bus. Rev.*, vol. 90, no. 10, pp. 60–68, 2012.

[20] P. Hitzler and K. Janowicz, "Linked Data, Big Data, and the 4th Paradigm," *Semantic Web*, vol. 4, no. 3, pp. 233–235, Jan. 2013.

[21] A. Zaslavsky, C. Perera, and D. Georgakopoulos, "Sensing as a Service and Big Data," *ArXiv13010159 Cs*, Jan. 2013.

[22] R. L. Villars, C. W. Olofson, and M. Eastwood, "Big data: What it is and why you should care," *White Pap. IDC*, 2011.

[23] 2013, "This recycling bin is following you," *Quartz*. .

[24] U. Greveler, B. Justus, and D. Loehr, "Multimedia content identification through smart meter power usage profiles," *Comput. Priv. Data Prot.*, 2012.

[25] G. Greenwald, "XKeyscore: NSA tool collects 'nearly everything a user does on the internet,'" *The Guardian*, 31-Jul-2013.

[26] J. Bamford, "The NSA is building the country's biggest spy center (Watch What You Say)," *Wired March*, vol. 15, 2012.

[27] "Twitter Usage Statistics - Internet Live Stats." [Online]. Available: http://www.internetlivestats.com/twitter-statistics/. [Accessed: 13-Jun-2014].

[28] T. H. Davenport, P. Barth, and R. Bean, "How 'big data'is different," *MIT Sloan Manag. Rev.*, vol. 54, no. 1, pp. 22–24, 2012.

[29] "Put the internet to work for you. - IFTTT," *IFTTT / Put the internet to work for you.* [Online]. Available: https://ifttt.com/. [Accessed: 11-Jun-2014].

[30] "Channels - IFTTT," *IFTTT / Put the internet to work for you.* [Online]. Available: https://ifttt.com/channels. [Accessed: 14-Jun-2014].

[31] "Too cold! Turn on the space heater," *IFTTT / Put the internet to work for you.* [Online]. Available: https://ifttt.com/recipes/141686-too-cold-turn-on-the-space-heater. [Accessed: 11-Jun-2014].

[32] C. C. Aggarwal and T. Abdelzaher, "Integrating Sensors and Social Networks," in *Social Network Data Analytics*, C. C. Aggarwal, Ed. Springer US, 2011, pp. 379–412.

[33] L. Atzori, A. Iera, and G. Morabito, "From 'smart objects' to 'social objects': The next evolutionary step of the internet of things," *IEEE Commun. Mag.*, vol. 52, no. 1, pp. 97–105, Jan. 2014.

[34] "Google now." [Online]. Available: http://www.google.com/landing/now/#whatisit. [Accessed: 15-Jun-2014].

[35] D. Bollier and C. M. Firestone, *The promise and peril of big data*. Aspen Institute, Communications and Society Program Washington, DC, USA, 2010.

[36] N. Eagle, A. (Sandy) Pentland, and D. Lazer, "Mobile Phone Data for Inferring Social Network Structure," in *Social Computing, Behavioral Modeling, and Prediction*, H. Liu, J. J. Salerno, and M. J. Young, Eds. Springer US, 2008, pp. 79–88.

[37] M. Aazam, I. Khan, A. A. Alsaffar, and E.-N. Huh, "Cloud of Things: Integrating Internet of Things and cloud computing and the issues involved," in *2014 11th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 2014, pp. 414–419.

[38] S. Distefano, G. Merlino, and A. Puliafito, "Enabling the Cloud of Things," in *2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, 2012, pp. 858–863.

[39] W. Wallace, *The logic of science in sociology*. Chicago: Aldine·Atherton, 1971.

[40] R. Minch, "Issues in the Development of Location Privacy Theory," in *Proceedings*, Kauai, Hawaii, 2011.

[41] P. Reynolds, *A primer in theory construction.* Indianapolis: Bobbs-Merrill, 1971.

[42] E. Nagel, *The structure of science problems in the logic of scientific explanation.* New York: Harcourt Brace & World, 1961.

[43] R. Handfield, "The scientific theory-building process: a primer using the case of TQM," *J. Oper. Manag.*, vol. 16, no. 4, p. 321, 1998.

[44] P. J. Phillips and A. J. O'Toole, "Comparison of human and computer performance across face recognition experiments," *Image Vis. Comput.*, vol. 32, no. 1, pp. 74–85, Jan. 2014.

[45] "Apple's iBeacon promises to do for indoor spaces what GPS did for the outdoors." [Online]. Available: http://online.wsj.com/articles/apples-latest-offering-explores-the-great-indoors-1401655436. [Accessed: 02-Jun-2014].

[46] "Location and Maps Programming Guide: Region Monitoring and iBeacon." [Online]. Available: https://developer.apple.com/library/ios/documentation/userexperience/conceptual/LocationAwarenessPG/RegionMonitoring/RegionMonitoring.html. [Accessed: 13-Jun-2014].

[47] R. P. <randy_presuhn@bmc.com>, "Management Information Base (MIB) for the Simple Network Management Protocol (SNMP)." [Online]. Available: http://tools.ietf.org/html/rfc3418. [Accessed: 24-Aug-2014].

[48] D. Namiot and M. Sneps-Sneppe, "Geofence and Network Proximity," in *Internet of Things, Smart Spaces, and Next Generation Networking*, S. Balandin, S. Andreev, and Y. Koucheryavy, Eds. Springer Berlin Heidelberg, 2013, pp. 117–127.

[49] A. F. Weston, *Privacy and Freedom.* New York: Atheneum, 1967.

[50] A. Juels, "RFID security and privacy: a research survey," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 2, pp. 381–394, Feb. 2006.

[51] D. Anthony, T. Henderson, and D. Kotz, "Privacy in Location-Aware Computing Environments," *IEEE Pervasive Comput.*, vol. 6, no. 4, pp. 64–72, Oct. 2007.

[52] A. S. Y. Cheung, "Location privacy: The challenges of mobile service devices," *Comput. Law Secur. Rev.*, vol. 30, no. 1, pp. 41–54, Feb. 2014.

[53] L. Hutchinson, "iOS 8 to stymie trackers and marketers with MAC address randomization," *Ars Technica*, 09-Jun-2014. [Online]. Available: http://arstechnica.com/apple/2014/06/ios8-to-stymie-trackers-and-marketers-with-mac-address-randomization/. [Accessed: 14-Jun-2014].

[54] E. D'educrat, "The Powers That Beat: Hacker Finds Easy Way to Secretly Save SnapChat Pictures," *The Powers That Beat*, 22-Jan-2013. .

[55] D. Gray and D. Citron, "The Right to Quantitative Privacy," *Fac. Scholarsh.*, Jan. 2013.

[56] M. Langheinrich, "Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems," in *Proceedings of the 3rd International Conference on Ubiquitous Computing*, London, UK, UK, 2001, pp. 273–291.

[57] "FTC Recommends Congress Require the Data Broker Industry to be More Transparent and Give Consumers Greater Control Over Their Personal Information." [Online]. Available: http://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more. [Accessed: 14-Jun-2014].

[58] D. Streitfeld, "European Court Lets Users Erase Records on Web," *The New York Times*, 13-May-2014.

[59] J. M. B. News, "Halt called to London tracking bins," *BBC News*. [Online]. Available: http://www.bbc.co.uk/news/technology-23665490. [Accessed: 12-Jun-2014].

[60] C. S. Fischer, *America Calling: A Social History of the Telephone to 1940*. University of California Press, 1992.

[61] A. Juels, R. L. Rivest, and M. Szydlo, "The Blocker Tag: Selective Blocking of RFID Tags for Consumer Privacy," in *Proceedings of the 10th ACM Conference on Computer and Communications Security*, New York, NY, USA, 2003, pp. 103–111.

[62] "Google Glass users fight privacy fears," *CNN*. [Online]. Available: http://www.cnn.com/2013/12/10/tech/mobile/negative-google-glass-reactions/index.html. [Accessed: 14-Jun-2014].

[63] "4 Places You Can't Use Your Cell Phone… and Why," *Connected Rogers*. [Online]. Available: http://www.connectedrogers.ca/news/4-places-you-cant-use-your-cell-phone-and-why/. [Accessed: 14-Jun-2014].

[64] "Public Policy | EPCglobal | Products & Solutions | GS1 - The global language of business." [Online]. Available: http://www.gs1.org/epcglobal/public_policy. [Accessed: 15-Jun-2014].

Table 1: Internet of Things information flow

Phase of information flow	Example components/methods
Sensing	Triangulation Scene analysis Proximity Indirect inference
Identification	Unique identifier detection Facial recognition Vehicle license plate recognition
Storage	Object data Meta data
Processing	Self-contained inferencing Communication and matching Advanced pattern recognition and data analytics
Sharing	Intentional Unintentional
Use	Intentional Unintentional

Table 2: Internet of Things example privacy measures

Phase of information flow	Example technical privacy control	Example social privacy control	Example legal privacy control
Sensing	RF blocking wallets RFID blocker tags [61]	Socially acceptable uses for Google Glass [62]	Prohibition of cell phone and camera use at customs [63]
Identification	MAC address randomization in Apple iOS 8 [53]	Anonymous letters to newspaper editors or postings to online discussion forums	“Secret” ballots for voting
Storage	No physical storage Encryption Ephemeral storage	User social media privacy settings	Formal limits on amount and duration of stored data
Processing	Privacy-enhancing technologies: anonymizing, etc.	Vendor-customer terms of service	Restrictions of database matching
Sharing	Restriction or non-provision of communication facilities	User and application sharing settings	The “right to forget” Data broker restrictions
Use	No provision for input into applications	Accepted business practices and standards such as EPC guidelines [64]	Prohibition of discriminatory use