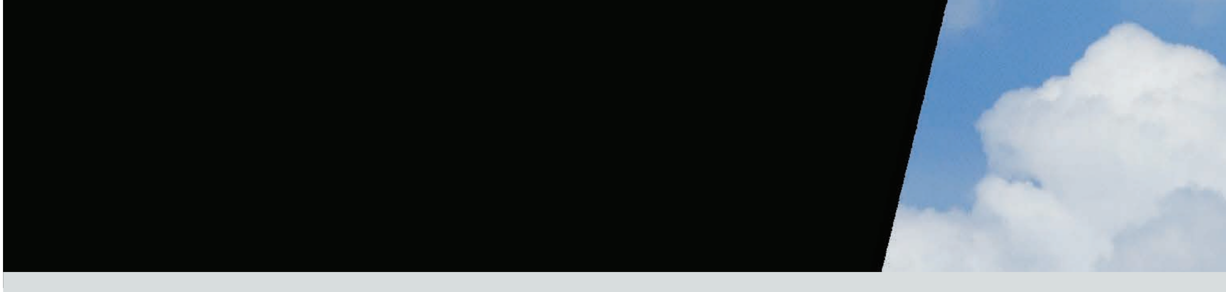


End-to-End Privacy for Open Big Data Markets

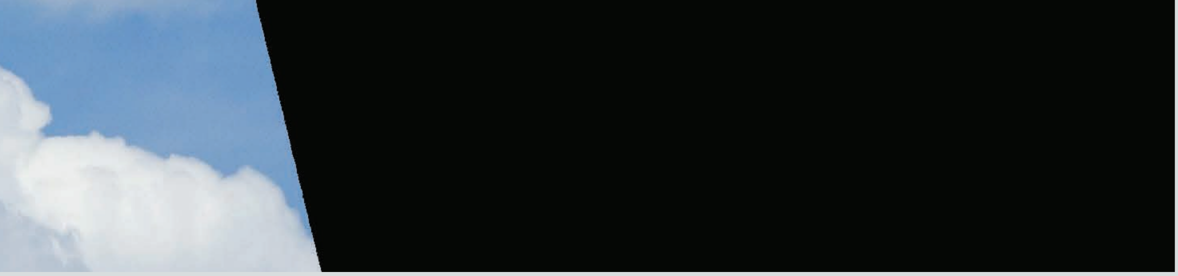
Charith Perera, Open University, UK
Rajiv Ranjan, Newcastle University
Lizhe Wang, Chinese Academy of Sciences

Data privacy is of great importance in the Internet of Things domain, especially in open data markets. This article discusses existing end-to-end privacy-preserving techniques and major research challenges.

The Internet of Things (IoT) promises to create a world where everyday objects (things) are connected to the Internet, either directly or through intermediate devices, and communicate with each other with minimum human intervention.[1] The ultimate goal is to create "a better world for human beings," where the objects around us know what we like, what we want, and what we need, and then act accordingly without explicit instructions. IoT lets people and things be connected anytime, anyplace, with anything and anyone, ideally using any path/network and any service.[2]

The current IoT marketplace clearly includes two broad categories of products and solutions.[3] Most of the products target individual customers (such as smart-home owners), who might expect comfort and convenience through some kind of automation. For example, WeMo is a Wi-Fi enabled switch that can be used to turn electronic devices on or off from anywhere.[3] Another example is Nest, a thermostat that learns what temperatures users like and builds a context-aware personalized schedule to automatically control the household temperature.[3] The second product group focuses on supporting business activities by collecting and analyzing sensor data in enterprise and industrial domains. The potential clients for these products are mostly companies, not individual customers. For example, SenseaAware supports real-time shipment tracking.[3] Context information such as location, temperature, light, relative humidity, and biometric pressure is collected and processed to enhance supply chain visibility. Another example, ParkSight, is a parking management technology designed for cities that

2325-6095/15/$31.00 © 2015 IEEE



retrieves context information through sensors (magnetometers) embedded in parking slots.[3]

Although the distinction between these two categories can sometimes be vague, we can identify some unique characteristics. The main unique characteristic is the target audience. In the first product category, potential clients are individual customers (that is, families). As a result, the data generated by the products should ideally belong to individual product owners. In contrast, the second product category targets enterprise customers. The data generated by this kind of solution might belong to the client company that bought the solution.

There are two important facts to highlight from the above discussion. First, it's important to understand that different IoT solutions capture different types of sensor data in different contexts (such as households, factories, or roads). Some IoT products might capture more privacy-sensitive information (for example, individual-customer-focused products) and others might capture less (for example, enterprise- or industry-focused products). The second important fact is that these IoT products typically focus on achieving a single objective and data always move within the solution boundaries. Therefore, because the data doesn't leave the product boundaries, the privacy risks related to these products are limited.

Despite these observations, a significant amount of useful knowledge and insight can always be derived by combining, processing, and analyzing the data collected by different IoT products.[4] For example, analysis of data collected by multiple data owners together can yield greater value than analyzing them separately. This type of data sharing approach is broadly referred to as *sensing as a service*.[4] The sensing-as-a-service business model is driving open big data markets. However, despite the potential value of such data sharing and knowledge discovery, such approaches can incur significant privacy risks. This article highlights the value of data sharing through open big data markets powered by the sensing-as-a-service model and provides design directions for ensuring end-to-end privacy.

Toward Liberated IoT Big Data

The sensing-as-a-service business model supports data exchange between data owners and data consumers.[4] Data owners purchase IoT products and deploy them in their own environments. These IoT products sense, analyze, and perform actuation to make the data owners' lives easier. As a by-product, the collected data is kept in access-restricted storage (usually referred as a *data silo*). Data consumers are entities that would like to access other peoples' data for some reason. For example, a data analyst in an energy company might want to know how many energy-inefficient legacy devices are used in a certain area. In this case, the data analyst isn't interested in a particular household, but a whole set of households. (We'll discuss different use case scenarios later in this article.) The presence of many data owners and potential data consumers creates an open data market. In this market, data might not be freely available for anyone to access; rather, only the metadata would be. Metadata would allow data consumers to understand what kind of data is stored in the silo. Interested data consumers must evaluate available metadata schemes and negotiate with the relevant data owners to gain access to their data. The sensing-as-a-service model primarily uses data generated by IoT products.

Data collected by different IoT products has a significant value when aggregated and processed on a large scale (for example, data collected from 10,000 households, where each house has 10 different IoT products). We discuss the details of sensing as a service elsewhere, and identify and analyze different types of data owners, consumers, and mediator service providers.[4] Although we haven't yet explicitly discussed the potential privacy issues, you can imagine how privacy violation could occur in this type of data sharing environment.

Motivation for End-to-End Privacy Protection

To understand the significance of privacy challenges in the IoT domain, it's important to visualize how each concept presented so far would work in the real world. Figure 1 illustrates the use case.

Let's introduce an example to help with our discussion. Jane is a restaurant manager who works different shifts. She lives alone in her own house. She has purchased (and deployed) three different IoT products in her house. The first is a context-aware thermostat that controls indoor temperature based on user preferences. She also has a smart

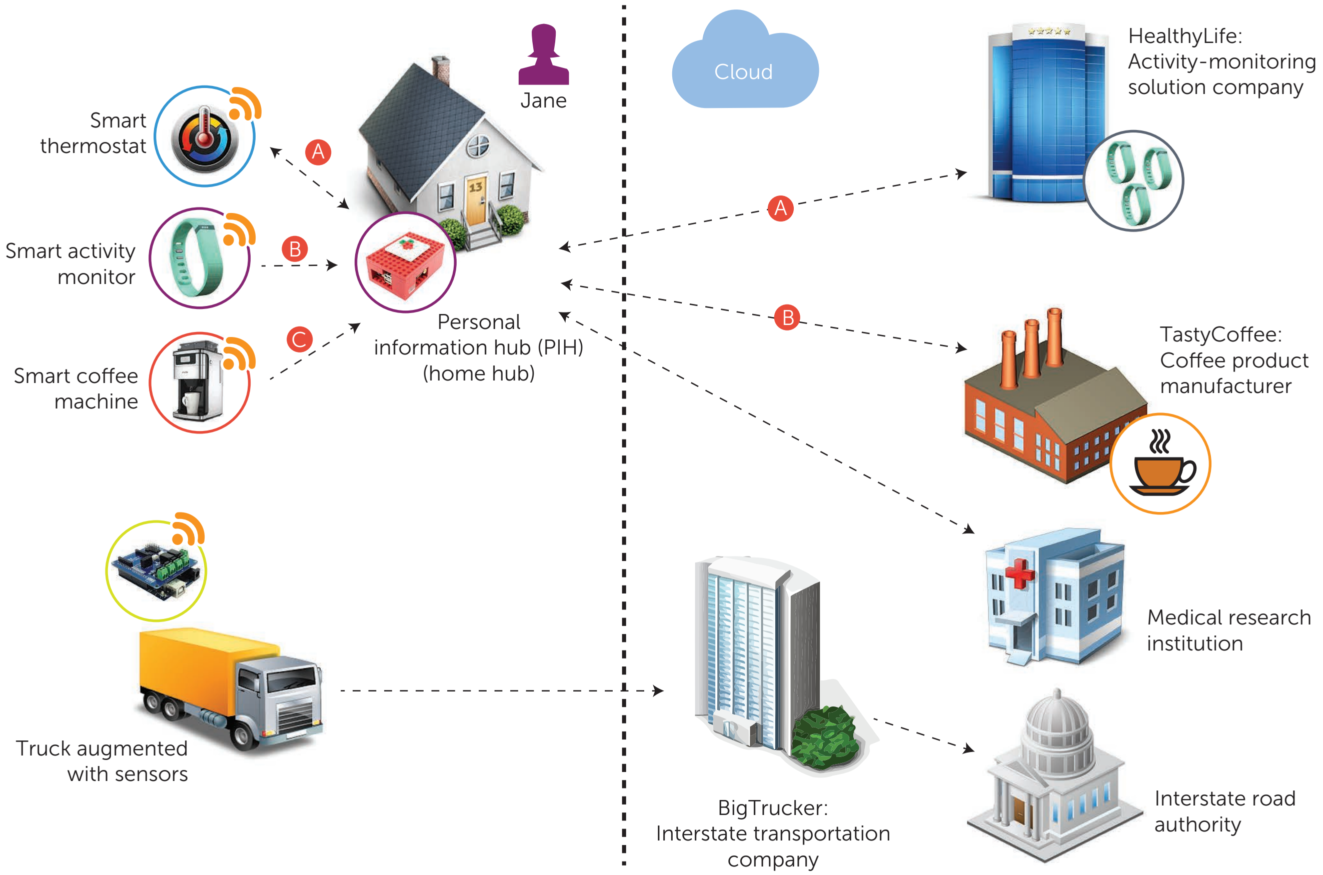

Smart thermostat
Smart activity monitor
Smart coffee machine
A
B
C
Jane
Personal information hub (PIH) (home hub)
Cloud
A
B
HealthyLife: Activity-monitoring solution company
TastyCoffee: Coffee product manufacturer
Medical research institution
Truck augmented with sensors
BigTrucker: Interstate transportation company
Interstate road authority

FIGURE 1. Open data market supported by the sensing-as-a-service model. Internet of Things-generated open data is ambient in every sphere of our lives, including smart logistics, smart homes, and remote healthcare.

coffee machine that automatically switches on and brews coffee when she gets up in the morning so by the time she arrives in the kitchen, coffee is ready for her. The third product is a smart activity monitor that tracks her exercise patterns, food intake, step counts, goals, and so on. Jane purchased and deployed these three products separately, and they work independently.

Data could move within these IoT solutions in different ways, depending on their functionalities and user requirements. Consider IoT products such as smart thermostats. These products learn user preferences over time and attempt to automatically actuate the heaters to control temperature. For this kind of actuation, the data collected by the product doesn't need to leave the house itself. Therefore, a small computer system built into the product (or using a Home Hub[3]) can process the data. These products use their own sensors to sense the environment and process the data within the household. Then, they actuate the actuators to perform certain tasks. We denote this type of dataflow in Figure 1 as *A*.

We can illustrate another type of dataflow using activity-monitoring health kits. These IoT products use their sensors to sense the environment and perform a certain amount of processing and actuating (for example, visualization and presentation, or notification). However, for further processing, some part of the data will need to be sent to the cloud services maintained by the product manufacturer. The purpose and advantage of such dataflow is that IoT product manufacturers can process data retrieved from a large number of users and give useful insights to the product owners in return. For example, if the data stays local, Jane will only be able to learn about her past, present, and future results based on her own data, which might not be very useful. However, if Jane lets her data be shared with the product manufacturer's service, she can compare her performance to similar users (for example, same age, weight, height, job, or workout patterns). Because the IoT product manufacturers access data from a large number of users, they can build more accurate, holistic, and comprehensive prediction models to support not only Jane but also others. This will create a community of users who share mutually beneficial IoT generated data. Jane would receive a benefit in return (money, a coupon, points on a shopping card, and so on) for giving her data to the IoT product manufacturer. However, at the same time, such dataflows involve potential privacy risks. We denote this type of dataflow in Figure 1 as *B*.

In the sensing-as-a-service model, another type of dataflow lets data owners, like Jane, give access to their data to a third party other than the respective IoT product manufacturer. We denote this type of dataflow in Figure 1 as *C*. Figure 1 shows TastyCoffee, a manufacturer of coffee products that's keen to know how people like Jane consume coffee (such as patterns and amounts). The company wants to know whether any external factors influence coffee consumption, such as weather, temperature, and workout patterns. For example, TastyCoffee would like to discover any consumer patterns (such as whether people tend to drink coffee before a workout). Currently, the only way the company can discover this type of information is through user surveys and focus group studies. However, such methods are time consuming, often inaccurate, and expensive. If TastyCoffee can access Jane's silo (along with those of thousands of similar users), which consists of data recorded from all three of her IoT products (smart thermostat, smart coffee machine, and activity-monitoring products), it will be able to understand Jane's activities (and those of thousands of similar users) better and optimize its product supply chain. Such optimization will allow TastyCoffee to reduce its costs and wastage, which would increase the company's profits.

Further, such data will help TastyCoffee improve its product lines and introduce new products to the market rapidly, which will also strengthen its brand value. Because of the additional value TastyCoffee might generate, it can offer a reward to the data owners to motivate them to give access to their data. From Jane's perspective, the additional reward would motivate her to trade her data not only with TastyCoffee, but also with other interested parties. This type of data trading creates more privacy risks than the two methods presented earlier.

In the TastyCoffee scenario, the data will be traded based on commercial interests. However, data trading in the sensing-as-a-service model could occur in a nonprofit way as well. For example, a medical research facility might be interested in accessing the same data as TastyCoffee, but with the intention of conducting research into people's well-being by analyzing the correlations between coffee consumption, exercise patterns, weather, and indoor temperature. In this scenario, the medical research center wouldn't be able to produce any direct financial profit, but it could use the research results to come up with actionable advice. For example, it might advise that consuming more than four cups of coffee reduces the impact of exercise by 20 percent (note that this is an entirely made-up fact we use to illustrate how an actionable advice might look; it isn't medical advice based on any scientific results), and pass along this advice to the data owners as a return.

Consider another example involving IoT products that we initially categorized as enterprise and industrial solutions. BigTrucker is a distribution company that handles goods on behalf of its clients (for example, it transports their goods between states). The company's trucks are augmented with sensors, which periodically sense the environment and report back to the BigTrucker management center. BigTrucker uses this IoT solution to monitor employees' health (such as work conditions over time), vehicle status (such as maintenance estimation), and the quality of the goods transported. However, interstate road authorities might be interested in accessing this data to understand environmental pollution and road conditions. Such data could help the authorities understand any environmental issues or infrastructure maintenance issues that need to be addressed urgently. Instead of deploying their own sensor networks and installing solar-based power supplies, authorities might request data from BigTrucker. In return, BigTrucker might receive financial compensation. In this scenario, data is traded between two parties, but the privacy risks involved are low because of the data's public and industrial nature.

As indicated by these scenarios, technology's responsibility is to support data trading in open data markets while protecting the privacy of all stakeholders. This is a main technological challenge we face today. In the remainder of this article, we survey existing privacy-preserving strategies and design techniques that can be used to facilitate end-to-end privacy for open IoT big data markets.

Technologies for Privacy Preservation

So far we've discussed why data trading between parties is important and how such activities can create significant value to all the stakeholders involved. At the same time, we implicitly highlighted why the privacy risk involved in such data trading is high. Here, we discuss how we can ensure that stakeholder privacy is protected when trading data by using existing privacy-preserving strategies and design techniques.

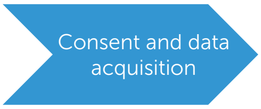


FIGURE 2. Different phases in a data life cycle including acquisition, transfer, storage, processing, and distribution.

Definition of Privacy

Before surveying privacy protection strategies and design technique in details, let's briefly define "privacy." Privacy is a concept in disarray, and it's difficult to articulate. "Privacy is far too vague a concept to guide adjudication and lawmaking, as abstract incantations of the importance of 'privacy' do not fare well when pitted against more concretely stated countervailing interests."[5] One widely accepted definition, presented by Alan Westin, describes information 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."[6] Roger Clarke has mentioned that "privacy is the interest that individuals have in sustaining a 'personal space,' free from interference by other people and organisations."[7]

Sometimes privacy is explained with the help of different dimensions. Privacy of the person, privacy of personal behavior, privacy of personal communications, and privacy of personal data are the four main dimensions of privacy.[7] The *Oxford Dictionary* defines privacy as "a state in which one is not observed or disturbed by other people" (www.oxforddictionaries.com/definition/english/privacy). More importantly, both the European Convention and the Universal Declaration of Human Rights have identified privacy as a human right. Further, the Charter of Fundamental Rights of the European Union defines the "respect for private and family life" in its Article 7 and adds a specific article on "protection of personal data" in Article 8. Additionally, Article 12 of the Universal Declaration of Human Rights protects an individual from "arbitrary interference with his privacy, family, home or correspondence," and "attacks upon his honour and reputation" (www.un.org/en/documents/udhr). This evidence strongly justifies the need to protect user privacy while we're attempting to harness the power of data trading and knowledge discovery to generate stakeholder value.

In parallel to the security protection goals, three goals have been proposed as primary privacy protection goals: unlinkability, transparency, and intervenability.[8] *Unlinkability* explains that data from multiple data sources shouldn't be combined in such a way that together they would violate user privacy. *Transparency* means that stakeholders must be informed about the data life cycle and what happens to each data item over time. This can be achieved through both technical and nontechnical means such as auditing, laws, and regulations. Data owners should know what types of data will be accessed, what types of data sources will be combined, where the data will be processed, what kinds of analytics will be used, what kinds of results will be generated, and so on. A step forward, *intervenability* says that data owners should be able to intervene at any time during the data life cycle so they can withdraw or change their consent at any time. More importantly, data owners should have control over their data all the time.

Phases in the Data Life Cycle

During the life cycle, data moves through different phases, as illustrated in Figure 2. Note that these phases are somewhat vague in the real world and the order could change depending on the context. Today, IoT data processing is moving from cloud computing to fog computing. The fog computing paradigm extends cloud computing and services to the edge of the network.[9] Similar to the cloud, fog provides data, computation, storage, and application services to users. The distinguishing fog characteristics are proximity to users, dense geographical distribution, and support for mobility. Processing data at the edge device avoids data communication and networking costs. Further, fog computing could reduce the potential for privacy violation (by, for example, processing smart-home data within the house itself). However, edge devices might have limited computational capacity, limited energy, and, more importantly, limited data and knowledge about a given context. To derive more insightful and useful knowledge, data might need to be combined and processed together. Therefore, in IoT, data processing location is a balancing act.

We've grouped some commonly used devices in the IoT domain into a few categories, as Figure 3 illustrates. This isn't a formal categorization based on any strict criteria. However, it approximates the differences between groups in terms of device capabilities. The devices belonging to each category

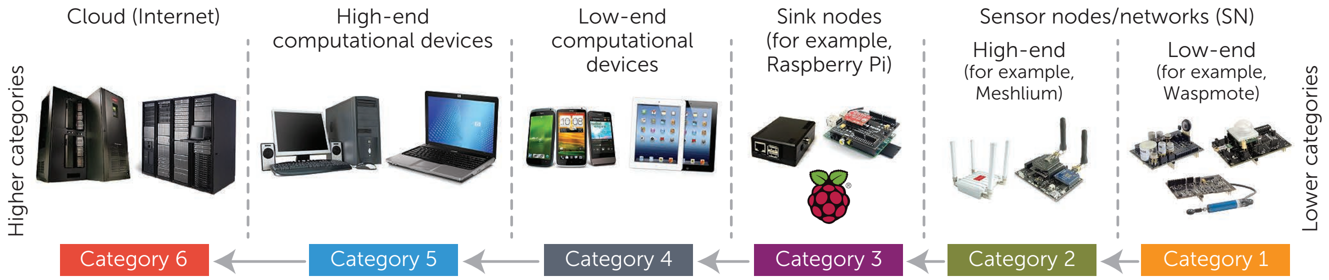

Cloud (Internet)
High-end computational devices
Low-end computational devices
Sink nodes (for example, Raspberry Pi)
Sensor nodes/networks (SN)
High-end (for example, Meshlium)
Low-end (for example, Waspmote)
Higher categories
Lower categories
Category 6
Category 5
Category 4
Category 3
Category 2
Category 1

FIGURE 3. Internet of Things devices can be categorized into six groups based on their computational capabilities. These category groups include clouds, high- and low-end computational devices, sink nodes, sensor nodes, and networking devices.

have different capabilities depending on processing, memory, and communication. They also differ in price, with the more expensive devices toward the left side of the figure. Computational capabilities also increase toward the left. Category 6 represents cloud computing; the other categories might act as edge devices depending on the context.

As might now be apparent, data transfer, storage, and data processing could occur iteratively as the data moves from right to left. However, the technologies behind these phases would remain mostly the same. Therefore, we combine them into the above-mentioned phases, even though their actual execution sequences can vary depending on the fog network's formation in a given context.

Privacy-Preserving Strategies and Design Techniques

Jaap-Henk Hoepman proposed several privacy-preserving strategies and design techniques.[10] These techniques can be applied to protect the IoT-generated data. Here, we briefly introduce those strategies from an IoT perspective, referring to different situations.

The *minimize* design strategy says that data consumers should only ask for the minimum amount of data they require to achieve their objective.[11] Typically, when data consumers ask for more data, it creates more risk for the data owners. As a result, data owners might be reluctant to trade their data. Additionally, data owners might expect a higher reward to match the additional risk incurred. This design strategy comes into play in the consent and data acquisition phase. In the sensing-as-a-service domain, negotiation will need to take place to reduce the amount of data that's being traded between parties by considering the associated risk and rewards. For example, if TastyCoffee wants to identify any pattern of coffee consumption and weather, it shouldn't request any data related to motion sensors deployed in Jane's house. The smart coffee machine can communicate with motion sensors to identify whether Jane is awake. However, such information has no value to TastyCoffee. Further, anonymization (such as removing identity information) and use of pseudonyms (removing identity and introducing the individual as a resident of Milton Keynes, for example) can also be used to minimize the amount of information traded.[12]

A pseudonym is an identifier of a subject used in place of the subject's real name. *Onion routing* enables anonymous communication over a network.[13] The sender remains anonymous because each intermediary node knows only the location of the immediately preceding and following nodes. This technique can be used to perform anonymizing aggregation over a large number of households. Instead of requesting data from a large number of households and aggregating it in a centralized location, onion techniques can be used to anonymously aggregate data on the fly.

The *informal* design strategy recommends embracing transparency and openness. This strategy is also relevant to the consent and data acquisition phase. However, it requires information about other phases to build a profile for both data owners and data consumers. Profiling is one of the most important tasks in open data markets because it supports data trading negotiations. Data owners should be informed about which data is processed, for what purpose, and by what means. It's important to let data owners know how the information is protected, and to be transparent about the system's security. This information will directly impact the data owner's preferences to trade with a particular data consumer. Because risk and reward are involved, trust plays a vital role in negotiating trades between data owners and data consumers. Approaches such as the Platform for Privacy Preferences (P3P, www.w3.org/P3P) can be used to model the data owners' privacy preferences, which could include their expectations about potential data consumers and their characteristics (such as level of trust, security, and openness about the techniques used in different phases of the life cycle).

The *hide* design strategy recommends hiding data from plain view. This strategy is useful in both data transfer and data storage phases. Different types of encryption techniques can be used during the data transmission from edge devices to cloud devices.[13] Data can be stored in different types of devices along the way as necessary. The encryptions supported by each device could vary depending on the device's computational capabilities. Today, encryption techniques are typically employed in data transfer and data storage phases. However, homomorphic encryption techniques have recently been introduced as a potential method to conduct computations over encrypted data.[14] When homomorphic encryption is used, data doesn't need decryption to be processed. Homomorphic encryption techniques can be incorporated with onion routing to support end-to-end privacy and security. For example, individual data silos could generate results based on the data consumers' requests, and the result would be passed from one silo to another, where each silo can append its results to the incoming result using homomorphic encryption. In this way, each silo would know its own results but would have no knowledge about the incoming data.

Today, encryption techniques are typically employed in data transfer and data storage phases.

The *separate* strategy recommends storing data in a distributed manner. In the IoT, this is the default assumption. Data owners may store their data in personal silos where they'll grant access to data consumers as a part of the trading process. This strategy is mostly related to the data storage phase but is also relevant to the data processing phase. There has been substantial research on distributed data storage, some are referred to as *personal information hubs* (PIHs). Examples include Hub of All Things (http://hubofallthings.com) and Lab of Things (www.lab-of-things.com). These edge devices sit inside the data owner's home. Broadly, PIHs can handle data processing using one of two methods. In the first, the PIH doesn't allow data to move outside its physical boundaries; Dataware is an example of this approach.[15] This approach employs a data analytical component into the PIH and allow it to perform data processing tasks within the PIH boundaries. Only the result will be sent out from the PIH. In the other method, data is considered movable and a limited amount of raw data will be sent out of the PIH. Data can then move to other silos or to the centralized cloud over the fog network, where data can be processed.

Another design strategy, *aggregate*, is more related to the data processing phase. This strategy recommends the release of only the aggregated results from data silos. Typically, data becomes less sensitive if it's sufficiently coarse grained, and the size of the group over which it's aggregated is sufficiently large. There are several ways to aggregate data, including aggregation within the PIH. In our previous example, instead of returning raw data to the data consumers, the PIH might return results saying that the data owner has used the coffee machine five times per day on average over the past three months (that is, aggregate over time). Such aggregated results don't provide detailed information about the coffee machine's usage. Another aggregation method is based on location. A potential result after distributed processing of multiple PIHs is, "40 percent of Milton Keynes households use energy-inefficient microwaves." Aggregation is tricky. For example, too much aggregation could hinder the knowledge-discovery process and data consumers won't be able to derive useful knowledge. On the other hand, trading less aggregated data could be too risky for data owners, as data consumers would be able to derive sensitive information about user behaviors and work patterns. Therefore, balancing the ideal level of aggregation is a challenging task. Techniques widely used in privacy-preserved aggregation are k-anonymity[16] and differential privacy.[17]

The *control* design strategy suggests that data owners should have the rights and access to the necessary tools to manage the data they trade to the data consumers. Again, this strategy is tricky because once data owners release results, it might not be possible to facilitate control functionalities that allow them to alter or remove their released data (that is, the results). Therefore, control in the IoT domain would be much more limited than privacy protection in traditional banking or healthcare domains. Specifically, if the PIHs are releasing aggregated and processed data, facilitating control would be impossible. However, control strategy is significantly valid in the early phases, when the data owner can decide which data to trade to which data consumers under what circumstances, and so on. Further, even after the data trading negotiations are done and contracts are in place, data owners should be able to change or cancel the contracts at any time.

The other two design strategies—enforce and demonstrate—are mostly nontechnical in nature and would potentially cover all different phases of the data life cycle. *Enforce* recommends that privacy policies be compatible with legal requirements. *Demonstrate* recommends establishing a data controller to demonstrate compliance with the privacy policy and any applicable legal requirements. This controller should be an independent third-party organization that can examine a given technology system (such as a given data consumer) and evaluate, audit, and log its behavior and level of compliance with privacy policies.

Research Challenges and Future Direction

Although there are many research challenges in privacy-preserving data analysis in the IoT domain, we focus on three major challenges that must be addressed before realizing the open data market vision.

Next-Generation IoT Middleware for Data Analysis

Since the 1990s, several guidelines have been proposed for designing and developing privacy-preserving software systems. Ann Cavoukian, for example, developed Privacy by Design to address the ever-growing and systemic effects of information and communication technologies, and large-scale networked data systems.[18] Although these design principles aren't specifically designed for the IoT domain, they encompass recommendations to build software systems that protect user privacy. Cavoukian proposed seven design principles: proactive not reactive (preventative not remedial), privacy as the default setting, privacy embedded into design, full functionality (positive-sum not zero-sum), end-to-end security (full lifecycle protection), visibility and transparency (keep it open), and respect for user privacy (keep it user-centric).

These design principles are still relevant in IoT domains as well. Further, the principles provide software designers, developers, and architects some direction on how to realize the vision of open data markets. In addition to the people who are directly involved in developing software, IoT envisions a strong community of data analysts who will be the force behind knowledge discovery. These people are in charge of deriving knowledge and insights from large volumes of data. In the sensing-as-a-service domain, they need to answer many questions on a daily basis—for example, what kind of data should be processed, what kind of analytics should be used, and where the data should be obtained. While answering such questions, they also need to make sure that user privacy is respected at all times. This is a challenging task, especially because data owners' privacy preferences and expectations vary. Further, accessing, transferring, storing, and processing data from each data source could require employment of a different privacy-preserving technique. It would be impossible for data analysts to handle such complexity manually. Therefore, we believe that there should be a middleware platform that allows data analysts to focus on data analysis and knowledge discovery tasks, while the middleware autonomously (or at least semi-autonomously) handles privacy-preserving techniques appropriately.

We've discussed various techniques that can be used to preserve user privacy during different phases of the data life cycle. It might already be clear that there are multiple methods to perform a given knowledge discovery task based on several factors (data movability, computational capability of edge devices, and so on). The IoT middleware platform should be able to autonomously combine different privacy-preserving techniques to support end-to-end privacy. Additionally, the middleware platform will need to help data analysts by providing useful tips (what kind of data is needed to discover certain knowledge or a particular pattern, what additional knowledge can be derived if more types of data are available, and so on) about which techniques to use if there's more than one way to accomplish a given task.

Conducting such composition tasks manually would be challenging, especially because of the numerous possibilities. For example, developers might write new data analytics components that could allow discovery of new knowledge. The ideal IoT middleware should be able to analyze these new data-analytical components and examine their potential impact on user privacy as well as where such components can be deployed (for example, on edge devices or in the cloud). Such IoT middleware would eliminate this significant burden on data analysts and reduce human error that could lead to user privacy violations.

Consent Acquisition and Negotiation

In the IoT, user consent involves acquiring the required level of permission from users and nonusers who are affected by the devices or services. In traditional Web agreements, user consent is received through the privacy terms and policies presented to users in paragraphs of text. With the recent emergence of social media and mobile apps, consent-acquiring mechanisms have changed. Researchers have found that current methods of requesting user consent in social media platforms, such as Facebook, are ineffective, and most users underestimate the authorization given to third-party applications.[19] In some cases, developers might not provide accurate information to users for the consent decision. In other cases, developers might provide accurate information, but users might be unable to understand exactly what the consent entails because they lack technical knowledge.

In the sensing-as-a-service domain, data owners are a major user type. Therefore, a major privacy challenge in the IoT, especially in relation to open data markets, is to develop technologies that request consent from data owners in an efficient and effective manner. This is a challenging task because every data owner has only limited time and limited technical knowledge to engage in the process. The consent-acquisition process is also part of the negotiation process. Research work in combining principles and techniques from human–computer interaction and the cognitive sciences are in dire need. Further, the sensing-as-a-service domain envisions that data consumers will request data from data owners. Sometimes, it would be difficult for data owners to spend much time evaluating these data requests. Therefore, there should be a way to build privacy profiles of individual data owner that encapsulate their privacy preferences. Such profiling can be achieved by questioning data owners about their privacy preferences combined with information about users' behavior and their data trading over time. When a data request is received, autonomous systems must evaluate the request on the data owner's behalf, performing a preliminary filtering to make the data owner's life easier.

Risk and Reward Modeling and Negotiation

After the preliminary filtering, the software systems on the PIH should provide the data owner with limited information, which might include risk and reward analysis in relation to a given data trading task. Data owners should have a complete picture of what's going to happen to their data and what they'll receive in return. Further, data owners should be able to negotiate with the data consumer regarding the amount of data to be traded and the related rewards. There are multiple ways to handle such negotiations, ranging from manual negotiations (that is, significant involvement of data owners) to autonomous negotiations. Data and consent acquisition should be a scalable process from both data owners' and data consumers' perspectives. Toward this end, semi-autonomous and autonomous negotiation strategies must be developed. Such strategies could consider factors such as data owners' preferences, how preferences have changed over time, and data consumers' requirements. Modeling privacy risks and conducting negotiations is challenging.[20]

Privacy protection isn't just an individual value, but also an essential element in the functioning of democratic societies. At the same time, open data markets that are expected to be created through the sensing-as-a-service model have a significant potential to generate value for the society by reducing wastage and costs, while allowing more personalized services to customers. A number of research gaps in the field need to be addressed to realize the vision of sensing as a service by creating open data markets. Future research efforts by the community will need to focus on addressing these research challenges.

Specifically, easy-to-use cloud-based privacy-preserving data analytics platforms will enhance the ability of data analysts to focus on data analysis tasks instead of worrying about privacy violations. Developing novel techniques to advise, recommend, and teach data owners about potential risks, threats, and rewards in the sensing-as-a-service domain will encourage more data owners to participate in open data trading. From a nontechnological viewpoint, incentive mechanisms in conjunction with strict auditing would help preserve user privacy while supporting useful knowledge discovery.

References

1. C. Perera et al., "Context Aware Computing for the Internet of Things: A Survey," *IEEE Comm. Surveys Tutorials*, vol. 16, no. 1, 2013, pp. 414–454.
2. H. Sundmaeker et al., *Vision and Challenges for Realising the Internet of Things*, Cluster of European Research Projects on the Internet of Things, European Commission, 2010.
3. C. Perera, C. Liu, and S. Jayawardena, "The Emerging Internet of Things Marketplace from an Industrial Perspective: A Survey," *IEEE Trans. Emerging Topics in Computing*, preprint, 8 Jan. 2015; doi: 10.1109/TETC.2015.2390034.
4. C. Perera et al., "Sensing as a Service Model for Smart Cities Supported by Internet of Things,"

Trans. Emerging Telecomm. Technologies (ETT), vol. 25, no. 1, 2014, pp. 81–93.

5. D.J. Solove, "A Taxonomy of Privacy," *Univ. of Pennsylvania Law Rev.*, vol. 154, no. 3, 2006, pp. 477–560.
6. A.F. Westin, *Privacy and Freedom*, The Bodley Head Ltd, 1967.
7. R. Clarke, *Introduction to Dataveillance and Information Privacy, and Definitions of Terms*, Xamax Consultancy Pty. Ltd., 21 Oct. 2013; www.rogerclarke.com/DV/Intro.html.
8. G. Danezis et al., *Privacy and Data Protection by Design—From Policy to Engineering*, tech. report, European Union Agency for Network and Information Security (ENISA), 2014.
9. F. Bonomi et al., "Fog Computing and Its Role in the Internet of Things," *Proc. 1st Ed. MCC Workshop on Mobile Cloud Computing*, 2012, pp. 13–16.
10. J.-H. Hoepman, "Privacy Design Strategies," *ICT Systems Security and Privacy Protection*, IFIP Advances in Information and Comm. Technology 428, N. Cuppens-Boulahia et al., eds., Springer, 2014, pp. 446–459.
11. S. Gürses, C. Troncoso, and C. Díaz., "Engineering Privacy by Design," *Proc. Computers, Privacy & Data Protection Conf.*, 2011.
12. A. Pfitzmann and M. Hansen, "Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management—A Consolidated Proposal for Terminology," 2010 https://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.31.pdf.
13. D. Goldschlag, M. Reed, and P. Syverson, "Onion Routing," *Comm. ACM*, vol. 42, no. 2, 1999, pp. 39–41.
14. C. Gentry, *A Fully Homomorphic Encryption Scheme*, doctoral dissertation, Dept. Computer Science, Stanford University, 2009.
15. D. McAuley, R. Mortier, and J. Goulding, "The Dataware Manifesto," *Proc. 3rd Int'l Conf. Comm. Systems and Networks* (COMSNETS), 2011, pp. 1–6.
16. L. Sweeney, "K-Anonymity: A Model for Protecting Privacy," *Int'l J. Uncertainty, Fuzziness, and Knowledge-Based Systems*, vol. 10, no. 5, 2002, pp. 557–570.
17. C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, nos. 3–4, 2013, pp. 211–407.
18. A. Cavoukian, "Privacy by Design: The 7 Foundational Principles: Implementation and Mapping of Fair Information Practices," Information & Privacy Commissioner, Ontario, Canada, 2009; www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/
19. J. Golbeck and M. L. Mauriello, *User Perception of Facebook App Data Access: A Comparison of Methods and Privacy Concerns*, tech. report, Human–Computer Interaction Lab, University of Maryland, 2014.
20. J.I. Hong et al., "Privacy Risk Models for Designing Privacy-Sensitive Ubiquitous Computing Systems," *Proc. 5th Conf. Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 2004, pp. 91–100.

CHARITH PERERA *is a research associate at the Open University, where he's currently working on the Adaptive Security and Privacy (ASAP) research program. His research interests include Internet of Things, sensing as a service, privacy, middleware platforms, sensing infrastructure, context awareness, semantic technologies, middleware, and mobile and pervasive computing. Perera has a PhD in computer science from the Australian National University, Canberra. He's a member of IEEE and ACM. Contact him at charith.perera@ieee.org.*

RAJIV RANJAN *is a senior research scientist and Julius Fellow at the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia, and an associate professor and reader in computing science at Newcastle University, UK. His research interests include cloud and big data computing, in particular quality-of-service-based management and processing of multimedia, Internet of Things, and big data analytics applications across multiple cloud datacenters (such as CSIRO Cloud, Amazon, and GoGrid); and automated decision support for migrating applications to datacenters. Contact him at rajiv.ranjan@csiro.au.*

LIZHE WANG *is a professor at the Institute of Remote Sensing & Digital Earth, Chinese Academy of Sciences, and a ChuTian Chair Professor at the School of Computer Science, China University of Geosciences. Wang has a PhD in engineering from the University Karlsruhe, Germany. He is a fellow of the Institution of Engineering and Technology and the British Computer Society, and a senior member of IEEE. Contact him at wanglz@radi.an.cn.*