Data Linkage in Smart Internet of Things Systems: A Consideration from a Privacy Perspective

Xu Zheng, Zhipeng Cai, and Yingshu Li

The authors investigate the privacy issues related to linkable data in smart IoT systems, which have not been thoroughly studied in previous works. They first discuss the available data sources in smart IoT systems and their linkage. Then they introduce some third parties who may access and utilize the linkable data. The potential threats are comprehensively elaborated for both individuals and crowds in smart IoT systems.

Abstract

Smart IoT systems can integrate knowledge from the surrounding environment, and they are critical components of the next-generation Internet. Such systems usually collect data from various dimensions via numerous devices, and the collected data are usually linkable. This means that they can be combined to derive abundant valuable knowledge. However, the collected data may also be accessed by malicious third parties to reveal sensitive information. In this article, we investigate the privacy issues of linkable data in smart IoT systems, which have not been thoroughly studied in previous works. We first discuss the available data sources in smart IoT systems and their linkage. Then we introduce some third parties who may access and utilize the linkable data. The potential threats are comprehensively elaborated for both individuals and crowds in smart IoT systems. Finally, we highlight some challenges and open problems for privacy preservation of linkable data in smart IoT systems.

Motivation

A smart Internet of Things (IoT) system primarily provides comprehensive, convenient, intelligent, and interactive services for both individuals and their surroundings. The corresponding techniques have been unprecedentedly developed and adopted due to the quick evolution of smart devices and the continuous investment of leading communities. Typical smart IoT systems include smart cities, smart grids, smart traffic, smart buildings, smart homes, and so on. A common feature of these systems is the support from data [1], which achieves fine-grained and comprehensive coverage for all participants. Nowadays, such data may include classical sensory data, contextual data, proactively uploaded contents, and interactions on devices. According to a recent study, the data collected by smart building systems alone will reach 37.2 zetabytes in 2020 [2]. In this article, we focus on privacy preservation of the data collected in smart IoT systems.

Smart IoT systems involve an increasingly huge number of users and service providers. All the users actively or passively contribute to a system with a variety of contents. A smart IoT system is able to provide typical services for all users as well as customized services for each individual user. This feature is quite different from the classical IoT systems, transforming an IoT system from a simple work cycle to a giant ecosystem.

Data linkage provides fundamental support for smart IoT systems. Data linkage means that the contents in a system can be combined to discover events. First, the contents collected from a single user are linkable. For example, sensory data and health data can be combined in a smart home system to intelligently control the temperature in a house. Second, the contents collected from different users are linkable. Take the smart meeting room as an example. The sensor on each chair can sense the presence of a user. Then the total number of participants in a meeting can be calculated by combining all the sensory data. Therefore, data linkage is quite critical for smart IoT systems, since isolated contents cannot provide sufficient information.

While bringing numerous benefits, data linkage also causes severe threats to user privacy. First of all, individual users may be more vulnerable, as more contents from more dimensions are collected. For example, users may not be aware of data collected by vehicles. If the collected data are linked to their smart home data, private information, such as daily activities, health status, and hobbies, may be disclosed. Furthermore, adversaries may infer the private information of a whole community. Contents published by multiple community members can be linked to derive the general information of a community. For example, a group of friends may expose their entire trip even if each of them only publishes some photos involving part of the trip. All such threats are accelerating user anxiety, and will severely thwart the adoption of smart IoT systems.

Unfortunately, no effort has been spent on designing privacy preserved smart IoT systems toward data linkage, although the number of successful attacks on smart home devices alone already exceeded 100,000 several years ago [3]. Current frameworks mainly consider the collected contents in smart IoT systems as isolated components [4], and the proposed mech-

Xu Zheng is with the University of Electronic Science and Technology of China;
Xu Zheng, Zhipeng Cai (corresponding author), and Yingshu Li are with Georgia State University.

Digital Object Identifier:
10.1109/MCOM.2018.1701245

0163-6804/18/$25.00 © 2018 IEEE

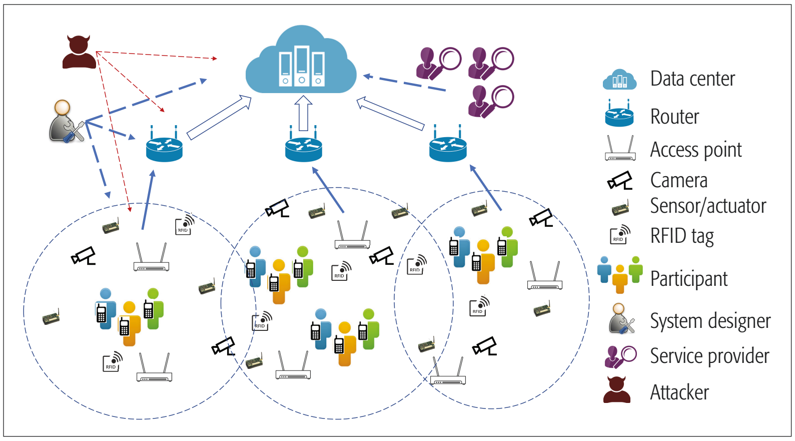

Figure 1. An example smart IoT system. The system designer deploys sensors, cameras, and so on, and also inspires mobile devices to collect linkable contents around participants. Then the contents are transmitted to servers via the hierarchical structure for storage and further processing. Finally, both system designers and service providers provide the corresponding services based on the collected contents. Attackers may also maliciously access or break into the system to filch the transmitted and stored contents.

anisms are for single dimensions [5]. They have also overlooked the heterogeneous behaviors of users and the correlations of the collected contents. Most works are based on simple access control [6] or local data sanitization [7]. Moreover, privacy preservation for a whole community has not been considered in smart IoT systems. To design a privacy preservation framework for smart IoT systems, data linkage should be the primary concern in all the procedures, including threat notification, access control, content publication, data sanitization, and so on, with the following objectives. The framework should help users better understand the privacy threats and numerous content dimensions. Dimension-based policies should be general and fine-grained. There should be a balance among utility, privacy, efficiency, and fairness. Data sanitization should be implemented according to the heterogeneous devices and their generated contents.

In this article, we investigate the privacy issues for linkable contents in smart IoT systems. We first introduce the common data sources in the current smart IoT systems. Then the underlying correlations among the collected contents are discussed. We also introduce the potential third parties that may access the linkable contents. Some threats for different scales of participants are presented as well. Finally, we address and highlight some challenges and open problems on the privacy preservation for current smart IoT systems. We use the terms data and content interchangeably as they both refer to collected information in systems.

Data Linkage

The pervasive emerging contents from numerous dimensions make smart IoT systems different from classical IoT. Linkable contents can be utilized to achieve comprehensive and reliable observations of all participants and events. In this section, we introduce the common data sources, the linkage among them, and the accessibility to the linkage.

Dimensions of Available Contents

In typical smart IoT systems [8], devices are deployed into a target area to collect contents from participants who could be all the residents in the area. In smart IoT systems like smart buildings and smart homes, third-party service providers are also allowed to define their own services. They cooperate with the system designers to refine the practicability of the whole system. We show an example smart IoT system in Fig. 1.

There are mainly three types of data sources in general smart IoT systems: sensory data from sensors, multimedia data from cameras and other monitoring devices, and actively and passively uploaded contents from mobile ends. They each generate contents in several dimensions like temperature, visualization, and reactions from participants.

Sensory data are collected via sensors, RFID tags, and so on. It is believed there will be more than 20 billion IoT devices installed by 2020 [9]. The densely deployed sensors can achieve fine-grained multi-dimensional coverage for the monitored area.

Multimedia data constitute a critical component for smart IoT systems. Smart IoT systems usually provide sophisticated services and request the support of contextual data. Therefore, images, videos, sounds, and other multimedia contents are necessary. As multimedia data usually carry abundant information, it is obvious that they can record the behaviors of multiple participants simultaneously.

Users of smart IoT systems actively participate in the systems. Participants can actively or passively upload contents from their mobile devices, including sensory data, images, and their interactions and reactions with the systems. These contents can help improve service quality, while also playing an essential role in customizing services since they are inherently linkable with the owners of the mobile devices.

Linkage of Contents in Smart IoT Systems

Considering the three types of data sources, there are mainly two types of linkage. The first one is the linkage among contents from multiple dimensions. The second one is the linkage among contents from different participants.

For the first type, the contents in different dimensions may be correlated with each other since they may record the same event at the same place from multiple aspects. Such a linkage is usually defined by the temporal, spatial, or contextual aspects. For example, contents from the sensors and cameras in the same building can be linked to track moving objects.

For the second type, contents from multiple participants in the same community or in a similar situation are linkable. In this case, the linkage means the contents may be used for the same purpose, or they record the same event or the general behaviors of a crowd. For example, family members may each submit contents in their own bedrooms, while these contents can be linked to derive the habits of the family. Such knowledge can also be used for services like temperature control.

Generally, both types of linkage are becoming more pervasive as smart IoT systems try to provide integrated coverage in the temporal-spatial-contextual space.

Techniques	Localization accuracy	Specifically designed for localization	Request for extra devices	Possibility for identification	Human awareness and controllability
GPS	Low	Yes	No	No	Easy
WiFi	Medium	No	No	Yes	Hard
Beacon	High	Yes	Yes	Yes	Hard
RFID	Very high	Yes	Yes	Yes	Easy
Camera	Medium	No	No	Partial	Easy
Sensor	Low	No	No	No	Hard

Table 1. Techniques for indoor localization: GPS is unstable due to the blockage in buildings; WiFi tracks the appearance in the nearby area; the beacon and RFID track the appearance in a smaller area than WiFi; the camera and sensor indirectly infer the appearance in monitored regions.

Accessibility of Linkable Contents

Considering the diverse dimensions of contents and their linkage, it is no wonder that the combination of these data sources can dramatically improve both diversity and quality of services. Current smart IoT systems also provide many opportunities to access such linkable contents, even for third parties ranging from dominating groups to unversed individuals who may be benign or malicious.

System designers are supposed to have access to all the contents generated in a system. They achieve this by simply collecting all the contents, sometimes even without a predefined purpose. Therefore, they are the ones who seize overwhelming control on all the linkable contents.

Third-party service providers may access linkable contents. They may not have full control over a system. However, they can implement different services in a system, and each service may access some linkable contents. Meanwhile, there are multiple ways to link contents. For example, each service can request identification of mobile devices. Then the service provider can link the contents from different services by combining their identifications. Furthermore, different service providers may cooperate and share obtained contents for more profits.

Data requestors can access linkable contents. Some smart IoT systems allow third parties to actively request contents from users. For example, the local government may request nearby individuals to collect emergency data in a target area. Then the data requestors can collect the contents from all the participants in the nearby area, which results in an inherent linkage in the time domain and spatial domain.

Malicious individuals or groups may access linkable contents. They achieve this by breaking into systems or acting as fake users and filching data.

Privacy Issues in Linkable Contents

While bringing fabulous benefits for better services in smart IoT systems, the linkage among contents provides more comprehensive coverage for individuals together with much side information, which may also severely breach the protection of sensitive information. Now we discuss the potential threats toward vulnerable individuals and the whole society under data linkage.

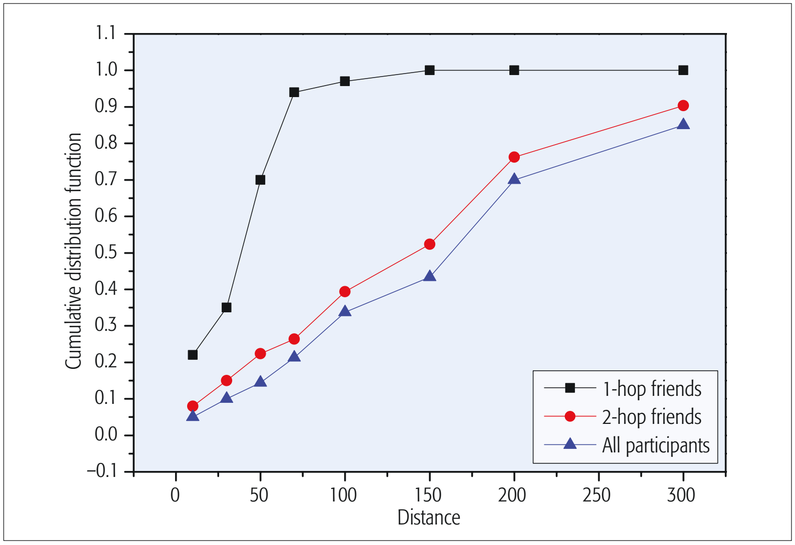

Figure 2. The distance between true destinations and predicted results. According to 1-hop friends, the errors could be approximately reduced to 100 km. However, the errors will increase when considering 2-hop friends. The reason is that some 2-hop friends may be weakly related to the target individual.

Individual Privacy in Content Dimensions

Various devices in smart IoT systems collect linkable contents from multiple dimensions. Once obtained by malicious third parties, the linked contents can be utilized to infer detailed information on the participants. First, they can be utilized to enhance the confidence of attackers. For example, the presence of an employee in an office can be simultaneously confirmed by her handheld devices, the sensors in the office, the usage of printers, and the door control system. More seriously, these attacks are hard to thwart as participants may have limited knowledge on the devices. They may have no idea about the strategy of the adversaries, which means the weak linkage among contents may also reveal sensitive information. Second, instead of inferring sensitive information, the linkable contents could be used to cover the full patterns of participants. In a smart building system, the comprehensive mobility pattern can be captured when all the contents are accessible, as there are multiple approaches to derive the location of a participant. Table 1 shows the categories of contents to derive the location of an individual. It includes nearly all types of devices in the system [10, 11]. More seri-

Some smart IoT systems like smart homes are designed for a group of related participants. As a result, the collected contents from multiple smart home appliances, the door control system, and the smart grids can be linked to achieve a fine-grained monitoring of the whole family.

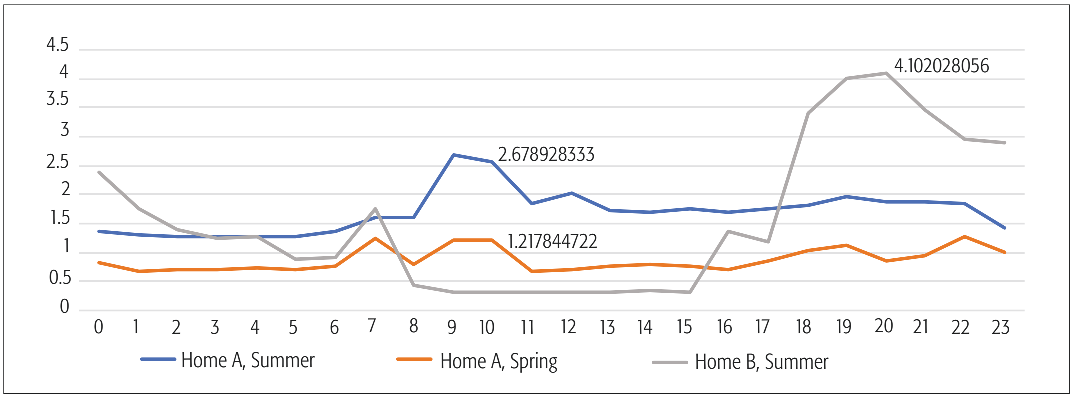

Figure 3. The electricity usage of two families: home A has relatively high usage in the morning between 9 a.m. and 11 a.m., while home B has a peak usage between 6 p.m. and 22 p.m.. Home A also maintains a smaller but same-shape usage in spring compared to the one in summer.

ously, attackers can derive the pattern even if they only have access to some of the contents (e.g., the sensory data). Generally, both threats come from the fact that the devices in smart IoT systems are inherently designed for cooperative and comprehensive monitoring of the environment. This indicates that the contents are definitely linkable and include useful information as well as sensitive information.

Individual Privacy in Communities

Due to the inherent correlations among participants, their published contents are usually linkable, which means these contents may include sensitive information for each other. This is the same as general online social networks where individuals may post photos including their friends [12]. We validate this threat via a public dataset recording the visited locations of individuals. The dataset is extracted from Gowalla [13], where participants post partial sequences of their visited locations together with their social relationships. The attackers utilize the Markov model to infer the missing sensitive locations, that is, the current location is only determined by its previous location. In Fig. 2, we can see that when the attackers know the social relationships of a participant, they can carry out more effective attacks by predicting the participant's destination according to the information released by friends. Actually, this kind of threat pervasively exists in smart IoT systems. The underlying reason is that participants usually work or stay with their colleagues and friends, and some of them may actively upload contents or be captured carelessly by nearby sensors or cameras. In this case, the sensitive information of a participant is vulnerable and uncontrollable due to the linkage of contents from her neighbors.

Community Privacy

Some smart IoT systems like smart homes are designed for a group of related participants. As a result, the collected contents from multiple smart home appliances, the door control system, and smart grids can be linked to achieve fine-grained monitoring of the whole family. On one hand, they can bring dramatic convenience to family members. On the other hand, they may reveal the private information of the family [14]. A comprehensive set of information, like the number of family members and their roles, their living patterns, and even their incomes, religions, and health status can be inferred. We take the electricity usage of two families as an instance [15]. The dataset records the average electricity usage of each family every 10 minutes. According to the usage shown in Fig. 3, we can see significant difference between the two families. This may indicate the heterogeneous habits of the two families. For example, home B could have more members since they have larger usage than home A, while the members in home A may stay at home during the day since they have high usage in the morning. If the usage data of other appliances are included, more sensitive information can be revealed. Another category of threats for a community come from smart IoT systems where community members actively upload contents via their mobile devices. Take a smart building system as an example. Suppose users from the same company can interact with the parking system, the vending machines, and the appliances in the dining room. While the contents are less sensitive for each participant, they can be linked to evaluate the working efficiency of a whole department.

Privacy for the Whole Population

Smart IoT systems may record information of a large enterprise or the whole population in an urban area. For example, a smart traffic system collects traffic data and predicts the migration of the population. If data brokers can access the contents, they can easily observe the events happening in the city and use the information for malicious purposes or commercial speculation. Furthermore, in some smart IoT systems, contents can be uploaded for rewards. Third parties are allowed to post their requests, and participants move to the requested locations to collect data for payments. Then malicious attackers like terrorists may utilize the systems to mislead the population. In both cases, the linkable contents capturing the whole population could lead to unprecedented coverage of public information. This coverage could lead to extremely dangerous occasions if obtained by malicious and dangerous third parties.

Generally, these privacy issues will bring severe threats to all levels of participants, and significantly limit the utility of the whole system.

Challenges	Subject	Adversary	Threats	Loss of quality	User interaction
Notification	System designers	Null	Null	No	Yes
Access control	System designers/service providers	Malicious service providers	Malicious access	No	No
Content publication: cooperative	System designers/ participants	Malicious data administrators	Over-collection	Partial	Yes
Content publication: local	Participants	Malicious data administrators	Individual inference attacks	Partial	Yes
Data sanitization	System designers	Data brokers	Data filching	Partial	Not

Table 2. Challenges for privacy preservation in smart IoTs.

As smart IoT systems collect data from numerous dimensions, unversed participants have little awareness of how the seemingly innocent physical contents could be related to their sensitive information. Therefore, there is obviously a barrier between the sensitive information beneath these information and user awareness.

Challenges and Open Problems

Considering the novel threats beneath the linkable contents, there are many challenges and open problems for the design of a privacy preservation framework, and a set of countermeasures are expected. We classify the challenges into four categories:
- User interaction
- Data access
- Data publication
- Data sanitization

The overview of the challenges is given in Table 2, which includes the information for the main components in each challenge.

Notification Regarding Latent Threats

As smart IoT systems collect data from numerous dimensions, unversed participants have little awareness of how the seemingly innocent physical contents could be related to their sensitive information. Therefore, there is obviously a barrier between the sensitive information beneath these data and user awareness. Then the challenge is how to properly notify the participants regarding the latent threats to their sensitive information.

However, it is usually infeasible to simply present the participants with a large scale of collected contents. For example, there are several approaches by which a smart building system can derive the location of a participant, and it may cover most regions in the whole building. In this case, simply showing the raw contents to participants is nearly useless. Therefore, it is essential to first derive the relationship between the collected contents and the events (i.e., localization) in the monitored area, and then study how the linkable contents could derive these events. Both tasks are challenging since how the environmental physical data relate to sophisticated events has not been thoroughly studied, and the consideration of linkable contents from different dimensions is still in the early stage. Furthermore, the heterogeneous contents involved in different IoT systems and the scale of the served participants are both quite diverse. These user-specific and system-specific features make the task even more challenging.

Even if the correlations are thoroughly derived, it is meaningless to present the unfamiliar raw contents to participants. Smart IoT systems should provide novel notifications to participants by informing them of the latent threats instead. The notification should be flexible, understandable, and comprehensive. First, participants should be able to subscribe to the pertinent contents or sensitive information so that accurate and meaningful results can be shown instead of continuously presenting raw contents. Second, the results must be understandable. For example, they should explain to what degree the concerned sensitive information is disclosed by which contents, or the percentage of the revealed sensitive information.

In summary, there should be a system-specific framework for the notification of sensitive information in smart IoT systems. It is expected to thoroughly consider data linkage and provide comprehensive and proper notifications for participants.

Access Control over Linkable Contents

In smart IoT systems, multiple service providers and even some skilled participants can access data and implement their services. Due to the pervasive existence of linkable contents, these roles can usually implement their services based on multiple combinations of contents. Therefore, malicious service providers or participants may request to include more sensitive information. All the facts aggravate the incapability of the existing naive dimension-based access control policies, and a more fine-grained authorization mechanism is necessary. Actually, there are some unique requests and challenges due to the linkage in smart IoT systems that make the design of the access control policies nontrivial. Below we introduce the challenges and open problems.

Fine-Grained: Instead of simply authorizing service providers to access contents in specific dimensions, there should be a more fine-grained policy that guarantees service providers access to just the contents necessary for their services. The underlying reason is that the accessed linkable contents may provide extra sensitive information. Then a more controllable framework is expected to help system designers determine the accessible contents. The main challenge is to derive the exact set of contents that can support the designed services while leaking minimum sensitive information. We call it function-based access control, which is usually nontrivial since there are usually multiple feasible content sets, and their correlations with sensitive information are diverse in both categories and seriousness. Meanwhile, these services usually request real-time access to contents, indicating that a smart IoT system should adaptively adjust the access policies for different service providers.

The access control policies for smart IoT systems should be built on thorough understanding of linkable contents, including their impacts on the targeted services and correlations with the concerned privacy. These policies should also consider the real-time cases where new contents keep emerging in a system.

Verifiable: Once authorized, service providers may access the contents and physical devices. They may abuse the authorization to collect unauthorized contents. Therefore, a second request is to regulate the access behaviors of service providers, that is, they can only acquire necessary contents even if they have more opportunities on controlling devices. To verify behaviors, there are two main challenges. The first one is to arrange and derive the rules for service providers. The rules should be clear enough such that smart IoT systems can use them to control the access on devices. The second challenge is to ensure that service providers follow the rules. Methods for cross-dimension tracking and auditing are required.

Collusion-Free: As a platform where multiple service providers can implement their services, a smart IoT system should properly control all contents accessed by these service providers. Otherwise, the service providers may share their independently innocent contents. For example, service providers for kitchen and light systems in a house may share their collected data, and achieve comprehensive eavesdropping on events in the family. The main challenge is to control privacy leakage to multiple service providers while retaining utility. Some other factors like efficiency are also supposed to be considered. This is quite challenging since collusions among service providers are unknown, and new service providers keep emerging. The first part indicates that over-protection will degrade service quality due to nonexistent collusions. The second part means that the access policy for subsequent service providers should consider previous ones. More specifically, they should get limited extra knowledge even if they share contents. Deriving the countermeasures is also difficult due to the diverse services and their heterogeneous requests on linkable contents.

Generally, the access control policies for smart IoT systems should be built on thorough understanding of linkable contents, including their impacts on the targeted services and correlations with the concerned privacy. These policies should also consider real-time cases where new contents keep emerging in a system.

Content Publication for Heterogeneous Participants

Participants should also preserve their own privacy. Generally, since participants are correlated, their linkable contents can be used for the same services. As a result, participants should properly choose their strategies for content publication. They need to preserve their own privacy as well as others' privacy, while retaining normal functions of smart IoT systems.

The first challenge is that participants are correlated by their contents, while their preferences on privacy and the contribution of their contents are heterogeneous. This fact requires published contents to be carefully decided. It must consider the privacy of each individual, while retaining good utility. To meet all these principles requires detailed knowledge on both utility and privacy of participants and their contents, which is difficult to attain. Meanwhile, other factors like resource consumption and fairness are also critical. Ideally, a smart IoT system should properly collect contents from public devices in public and personal places and personal devices, while maintaining fair privacy preservation among participants. As a consequence, several unsettled problems in this category include:

- The derivation of linkage between participants and their contents
- Exploring the leakage of one's sensitive information from her locally published contents and others'
- Cooperative content publication mechanisms considering utility, privacy, efficiency, and fairness

Second, for every single participant, there should be some mechanisms to locally regulate their behaviors in smart IoT systems. Unlike traditional IoT systems, smart IoT systems have far more devices for data collection as well as more available data sources. Both facts contribute to novel scenarios for participants. The challenges come in two parts. First, participants are confused and frustrated by so many notifications. It will be inefficient to keep notifying them when participants request most services. Second, participants are unfamiliar with the collected contents and the devices, which means that participants will find it difficult to follow the principles for behavior regulation due to the large number of rules and data linkage.

To properly handle both challenges, some open problems should be solved. First, there should be a framework to automatically learn user preference on privacy preservation. The framework should be able to recommend proper settings for different participants. Second, the guidelines for behavior regulations are unsettled. The rules should be able to help participants minimize the publication of unnecessary contents while retaining qualified user experience.

Data Sanitization Mechanisms

Finally, as devices in smart IoTs collect linkable contents, to conceal on a single device may be insufficient. For instance, obfuscating sensory data and uploaded contents may conceal sensitive information for participants. However, when two dimensions of contents are combined, the perturbation will be degraded and the latent sensitive information revealed. Therefore, the content obfuscation mechanisms for privacy preservation should be improved for smart IoT systems. Novel mechanisms should take data linkage into consideration and achieve global privacy preservation. It may share contents with nearby devices, and then determine the suitable strategies for data sanitization. The proposed solution must consider the extra cost introduced by sharing.

Conclusion

This article investigates the privacy issues in smart IoT systems that are introduced by data linkage among collected contents. More specifically, it validates how the linkage among multiple contents may reveal private information of participants. To achieve this goal, the article first introduces the abundant available data sources, and discusses the underlying linkage among these contents. Then some parties that may access the data and learn the linkage are introduced. A set of existing threats on private information are demonstrated. Finally, a comprehensive discussion on the open problems of privacy preservation in smart IoT systems is presented to address the design of novel notification mechanisms, access control policies, content publication strategies, and data sanitization mechanisms.

Acknowledgment

This work was conducted while the first author was a Ph.D. student in the Department of Computer Science, Georgia State University. This work is partly supported by the National Science Foundation (NSF) under grant nos. 1252292, 1741277, and 1704287; the National Natural Science Foundation of China (NSFC) under Grant nos. 61502116 and 61632010; and the Foundation of Science & Technology Department of Sichuan province (no. 2017JZ0031)

References

[1] Y. Sun *et al.*, "Internet of Things and Big Data Analytics for Smart and Connected Communities," *IEEE Access*, vol. 4, 2016, pp. 766–73.

[2] "Volume of Data Collected by Smart Buildings"; https://www.statista.com/statistics/631151/worldwide-data-collected by- smart-buildings/, accessed Nov. 7, 2017.

[3] V. Sivaraman *et al.*, "Network-Level Security and Privacy Control for Smart- Home IoT Devices," *Proc. 2015 11th IEEE Int'l. Conf. Wireless and Mobile Computing, Networking and Commun.*, 2015, pp. 163–67.

[4] R. H. Weber, "Internet of Things — New Security and Privacy Challenges," *Computer Law & Security Review*, vol. 26, no. 1, 2010, pp. 23–30.

[5] S. Jana, A. Narayanan, and V. Shmatikov, "A Scanner Darkly: Protecting User Privacy from Perceptual Applications," *Proc. 2013 IEEE Symp. Security and Privacy*, 2013, pp. 349–63.

[6] E. Fernandes *et al.*, "Flowfence: Practical Data Protection for Emerging IoT Application Frameworks," *Proc. USENIX Security Symp.*, 2016, pp. 531–48.

[7] X. Hu *et al.*, "Differential Privacy in Telco Big Data Platform," *Proc. VLDB Endowment*, vol. 8, no. 12, 2015, pp. 1692–1703.

[8] X. Cui, "The Internet of Things," *Ethical Ripples of Creativity and Innovation*, Springer, 2016, pp. 61–68.

[9] "Here's How the Internet of Things Will Explode by 2020"; http://www.businessinsider.com/iot-ecosystem-internet-of-thingsforecasts- and-business-opportunities-2016-2?IR=T, accessed Aug. 31, 2016.

[10] C. Xu *et al.*, "Towards Robust Device-Free Passive Localization through Automatic Camera-Assisted Recalibration," *Proc. 10th ACM Conf. Embedded Network Sensor Systems*, 2012, pp. 339–40.

[11] J. Xiao *et al.*, "A Survey on Wireless Indoor Localization from the Device Perspective," *ACM Computing Surveys*, vol. 49, no. 2, 2016, p. 25.

[12] N. Z. Gong and B. Liu, "You Are Who You Know and How You Behave: Attribute Inference Attacks via Users' Social Friends and Behaviors," *Proc. USENIX Security Symp.*, 2016, pp. 979–95.

[13] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and Mobility: User Movement in Location-Based Social Networks," *Proc. 17th ACM SIGKDD Int'l. Conf. Knowledge Discovery and Data Mining*, 2011, pp. 1082–90.

[14] D. Townsend, F. Knoefel, and R. Goubran, "Privacy Versus Autonomy: A Tradeoff Model for Smart Home Monitoring Technologies," *Proc. 2011 IEEE Annual Int'l. Conf. Engineering in Medicine and Biology Society*, 2011, pp. 4749–52.

[15] S. Barker et al., "Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes," *SustKDD*, Aug., vol. 111, 2012, p. 112.

Biographies

Xu Zheng (xzheng7@student.gsu.edu) received his Ph.D. and M.S. degrees from the School of Computer Science and Technology, Harbin Institute of Technology. He is currently a Ph.D. student in the Department of Computer Science, Georgia State University. He will join the School of Computer Science and Engineering at University of Electronic Science and Technology of China. His research areas focus on IoTs, smart cities, and data privacy.

Zhipeng Cai (zcai@gsu.edu) received his Ph.D. and M.S. degrees from the Department of Computing Science, University of Alberta, and his B.S. degree from the Department of Computer Science and Engineering, Beijing Institute of Technology. He is currently an associate professor in the Department of Computer Science, Georgia State University. His research areas focus on networking and big data. He is the recipient of an NSF CAREER Award.

Yingshu Li [SM] (yili@gsu.edu) received her Ph.D. and M.S. degrees from the Department of Computer Science and Engineering, University of Minnesota-Twin Cities. She is currently an associate professor in the Department of Computer Science, Georgia State University. Her research interests include wireless networking, sensor networks, sensory data management, social networks, and optimization. She is the recipient of an NSF CAREER Award.

When two dimensions of contents are combined, the perturbation will be degraded and the latent sensitive information is revealed. Therefore, the content obfuscation mechanisms for privacy preservation should be improved for smart IoT systems. Novel mechanisms should take data linkage into consideration, and achieve global privacy preservation.