Editors: Heather Richter Lipford, richter@uncc.edu | Jessica Staddon, jessica.staddon@gmail.com

The Last Mile for IoT Privacy

Richard Chow | Intel Corporation

According to Mark Weiser:[1]

> *The problem, while often couched in terms of privacy, is really one of control. If the computational system is invisible as well as extensive, it becomes hard to know what is controlling what, what is connected to what, where information is flowing, how it is being used ... and what are the consequences of any given action.*

Weiser was discussing ubiquitous computing more than a decade ago, but he might as well have been talking about the Internet of Things (IoT) today. On our computers, we have at least a semblance of control because we can, in principle, determine what applications are running and what data they're collecting. For the IoT, traditional methods of control are largely absent. In fact, there are common cases where people are no longer *users* of an IoT service but rather *subjects* of the service, such as a smart city sound monitor. Another example is public Wi-Fi connectivity. A fraction of people in the vicinity might discover and use the service, but not being the ones who installed the actual access points, most people would be unaware of the service's privacy properties.

One of the IoT's major privacy problems is that users aren't always aware when a device is collecting personal data. IoT devices' ubiquitous nature means that a person can easily not know when sensors are present. A basic privacy tenet, dating back to the Fair Information Practice Principles, states that personal data collection should happen only with appropriate notice. Therefore, one of the goals in this article is to provide a framework—called the privacy stack—for user communication issues and needs regarding IoT privacy (see Figure 1).

The Privacy Stack

Many design questions, such as discovery, usability, and privacy, involve end users. How do users learn about the services in public spaces, their workplaces, or even their own homes? How do they learn about these services' privacy properties and yet not be overwhelmed by the sheer volume of information? I give a couple motivating examples of IoT privacy notifications.

Suppose a mall installs a suite of surveillance cameras in its parking structures for security purposes. Rather than describing the cameras' purpose on physical signage, the mall uses beacons to interact with phones carried by passersby. The beacons emit an ID that can be looked up in the cloud using a standard mobile app. The users can thus be alerted to the surveillance cameras' presence, which not only serves as a privacy notification but also provides them with a sense of security knowing that the area is safe due to the service.

Privacy notifications can also be incorporated into smart home systems. Consider open space voice controllers such as the Amazon

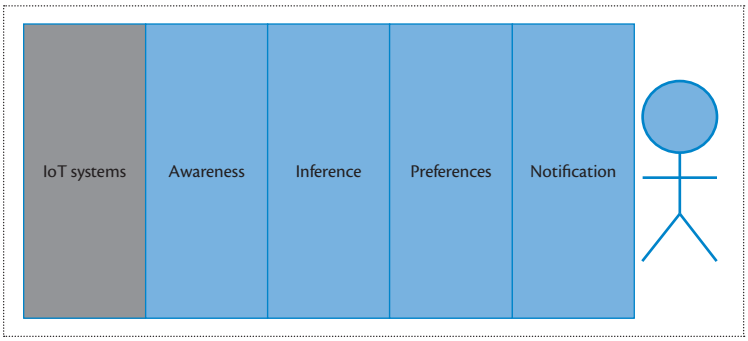

Figure 1. The privacy stack framework bridges from today's Internet of Things (IoT) systems to users.

Echo, which allow voice commands to play music, adjust lighting, and so on. Guests to the home might not be comfortable with a whole-home voice-recording system. Homeowners might ask permission to use such a service if their guests indicate that they'd like to be notified about audio recording in their vicinity. Depending on the guests' preferences, they could be notified as they enter the house, or alternatively, the system could disable the service temporarily or partially. All this might be done silently or explicitly.

In these motivating examples, the users don't understand what services are active in the environment. In the sections that follow, I describe how the examples might be realized through the privacy stack shown in Figure 1, which takes us from today's IoT protocols to human notification. I emphasize that this stack isn't a protocol stack, but simply a conceptual framework.

Awareness

The awareness part of the stack provides for discovery of services' privacy properties by users or users' agents. The difficulty with the IoT is the potentially unobtrusive nature of data collection. One can, in fact, not realize that data is being collected. According to Judith Donath, a fellow at Harvard's Berkman Center for Internet and Society, we must design so that we can know how public or private a space is, to know how to act and how candidly to speak.[2] Along these lines, Jason Hong has issued a challenge to the pervasive research community: can we make it so that when people enter a room, they can reliably identify all of the sensors and dataflows within 30 seconds?[3]

Awareness primarily concerns how IoT services might open communication channels to users and subjects. These channels might be new visual signifiers of data collection,[4] or more traditional network protocols between the IoT service and a user's device. For protocols, multiple industry efforts, including the Open Connectivity Foundation (openconnectivity.org), are working on standardizing IoT interoperability, allowing discovery and communication among devices.

Apple's iBeacon is another example of device discovery (developer.apple.com/ibeacon). A beacon device broadcasts an ID through Bluetooth, and a compatible device can use this ID to retrieve associated information (such as for location-based services). Beacons can also place a person in proximity to devices in the environment.

Thus far, the IoT protocol work has not gone into privacy metadata standardization, a way for services to declare their privacy policies so that they are universally understood. One advantage of doing so would be that, in a world of ubiquitous services, privacy decisions can be made with minimal cognitive burden. But even for the web, we don't have privacy metadata standardization for various reasons. One interesting approach that sidesteps standardization is to build natural language processing tools for privacy policies (for example, see www.signifiers.io). Nevertheless, the risks of data collection with completely unaware subjects are greater for the physical world and IoT, which argues for giving a standard language for IoT privacy another try.

Inferences

Awareness protocols can enable IoT services to declare what they're doing, but what do they declare? It seems simplest to declare what sensor data is collected and what it will be used for, but this might be insufficient. Users have limited understanding of what might be inferred or learned from sensor data, and systems will only get better at learning from sensor data. For instance, device or browser fingerprints are now a fact of life on the web, and it's reasonable to assume that fingerprints are even more prevalent in data measuring in the physical world. A couple of examples are location patterns and, more recently, ambient audio.[5] The inference problem is central to privacy, and yet it's unclear how to declare inferences; for instance, how does the system handle probabilistic inferences and inferences with auxiliary data?

I propose a balanced approach here: users can't be relied on to understand the inferences possible from the data collected, so services must explicitly provide basic inferences. At the same time, users must understand that this set of inferences is continually growing and refined, not only through the never-ending growth of user data but also through the advancement of machine learning techniques. One way for IoT services to provide

inferences is through the privacy policy; for instance, an IoT service collecting GPS data would declare that identity is a possible inference from the data, and a device collecting energy usage data would declare behavioral patterns as a possible inference.

An understanding of inferences wouldn't just help users understand what the system is learning about them but would also help systems protect privacy by translating user preferences. For instance, a system might collect video data, timestamps, and Wi-Fi data. A user might be comfortable with sharing video data collected at work but not video data collected at home. Security policies for the raw data itself can be implemented, but only after the system uses the data to infer "work" versus "home." This highlights the gap between low-level raw data and the language of human preferences, which consists of higher-level concepts such as "home." Sensors operate at a different level from what users find meaningful, and inferences help bridge the gap to usable systems.

Privacy Preferences

Moving on to the stack's next layer, suppose the system has a good understanding of what data an IoT service is collecting and what inferences might be possible. What does the user or subject actually care about? Whether a particular IoT data collection scenario is considered privacy sensitive depends on the individual, but context is also critical. For instance, audio recording might be fine in a restaurant depending on who's doing the recording (restaurant or friend) and who you're with (work or social). Context also gives clues as to what's surprising and what's expected. Video surveillance in a football stadium might be unremarkable, but perhaps not at a restaurant.

Because so many factors might influence a privacy decision, the central question is: what factors do people really care about in an IoT context? The sidebar provides a sampling of what the current research has found.

This research is just beginning, but the vision is that an understanding of the relevant factors will encourage system creators to present this information to those making privacy decisions. Knowing that particular factors might be important to some users can also influence system design. For instance, an IoT service at a bar might set patrons at ease by informing them that video is being captured by the bar for age-verification purposes only.

Notification

The stack's last layer involves notification of the end user. This notification depends on awareness of the services in the environment

Internet of Things Privacy Preferences: Current Research

What privacy factors are users most concerned about in an Internet of Things (IoT) context? Here's what some recent research has discovered.

Hosub Lee and Alfred Kobsa approached the problem by studying user reactions to scenarios with varying context elements.[1] For each scenario, they chose from a small set of parameter values in the following categories: who was collecting the data, where the data was being collected, what kind of data was being collected, the reason for collection, and the persistence of collection. An example scenario had "government" for who was collecting and "safety" for the reason for collecting. They found clusters in this context space that predicted user reaction.

Reuben Binns and his colleagues pointed out that privacy decision making is influenced by particular attributes, such as the reputation and size, of the company doing the collecting.[2]

Linda Lee and her colleagues investigated the "what" and "who" factors for wearable devices; the "what" had more effect than the "who" on user reactions, and users were most concerned about the collection of video or photos.[3]

Pardis Emami Naeini and her colleagues provided more statistical evidence for the impact of various factors, such as what and where, and also advanced explanations for the impact, such as perceived benefit.[4]

References

1. H. Lee and A. Kobsa, "Privacy Preference Modeling and Prediction in a Simulated Campuswide IoT Environment," *Proc. IEEE Int'l Conf. Pervasive Computing and Communications* (PerCom 17), 2017, pp. 276–285.
2. R. Binns et al., "My Bank Already Gets This Data: Exposure Minimisation and Company Relationships in Privacy Decision Making," *Proc. CHI Conf. Extended Abstracts on Human Factors in Computing Systems* (CHI EA 17), 2017, pp. 2403–2409; doi.acm.org/10.1145/3027063.3053255.
3. L. Lee et al., "Information Disclosure Concerns in the Age of Wearable Computing," *Proc. NDSS Workshop Usable Security* (USEC 16), 2016; dx.doi.org/10.14722/usec.2016.23006.
4. P. Emami Naeini et al., "Privacy Expectations and Preferences in an IoT World," *Proc. 13th Symp. Usable Privacy and Security* (SOUPS 17), 2017; www.usenix.org/conference/soups2017/technical-sessions/presentation/naeini.

and considers inferences and user preferences.

Notification is envisioned to be a relatively rare event, as users won't want to interact with most IoT services when they become ubiquitous; even services that merit a notification the first time might not the second or third time. For instance, drivers might want to be alerted when their car enters a region where it might be tracked but won't want to be alerted every day for the same street camera.

Central questions for notifications include: How does a system present alerts, how does a user express alert preferences, and how are these preferences learned? One notable point is that an alert that's not privately communicated could allow others to draw undesirable inferences about a user's history or preferences. Also, technology and inference algorithms might improve enough such that people who were previously notified about a service might need to be notified again.

Notification represents the last layer of the privacy stack: actual interaction with users. But other forms of interaction are possible. Notification implies interrupting a user involved in another activity, but a more active user might be interested in visualizing the privacy properties for nearby IoT services. How does a user visualize nearby IoT services and their privacy and trust properties?

Implementing the Stack

To illustrate, I describe how the stack might be implemented in the commonly proposed privacy proxy or privacy assistant on a user's device.[6,7] For awareness, the privacy proxy manages protocols that pull (or are pushed) privacy metadata from nearby IoT services. Basic inferences might be included in the privacy metadata, and the privacy proxy might also have the intelligence to make its own inferences from the data. The proxy manages personalized privacy preferences, for instance, allowing users to declare preferences, or inferring them from past privacy decisions or demographics. A privacy proxy builds and maintains a user's privacy preferences and decisions, and thus the proxy can use intelligence to alert the user only when appropriate.

A basic privacy principle is that personal data collection should happen only with appropriate notice and choice. However, this principle's implementation—already difficult for traditional clients—is even more difficult for the IoT because there's no natural communication channel with users. In many cases, users might find themselves in environments where services are running and capturing their data, even though the users haven't installed the services and aren't aware that they're running.

A conceptual privacy stack describes how IoT systems being developed today interact with users. This stack is being actively researched in multiple areas. One area is the user interface, both for setting and learning privacy preferences and for notification and visualization. A second area is the inference problem for IoT data—how to translate raw sensor data into human-understandable inferences. A third area of research is the privacy schema and how to represent the plethora of IoT services and their privacy policies, which is critical for automatic filtering. This is an exciting and active area, combining machine learning, human–computer interaction, and privacy. ■

References

1. M. Wieser et al., "The Origins of Ubiquitous Computing Research at PARC in the Late 1980s," *IBM Syst. J.*, vol. 38, no. 4, 1999, pp. 693–696.
2. "Five Minutes with Judith Donath," The MIT Press, 30 July 2014; mitpress.mit.edu/blog/five-minutes-judith-donath.
3. J. Hong, "The Privacy Landscape for Pervasive and Ubiquitous Computing," *IEEE Pervasive Computing*, vol. 16, no. 3, 2017, pp. 40–48.
4. S. Zimmeck et al., "Automated Analysis of Privacy Requirements for Mobile Apps," *Proc. 24th Network and Distributed System Security Symp.* (NDSS 17), 2017; dx.doi.org/10.14722/ndss.2017.23034.
5. N. Karapanos et al., "Sound-Proof: Usable Two-Factor Authentication Based on Ambient Sound," *Proc. 24th USENIX Conf. Security Symp.* (SEC 15), 2015, pp. 483–498.
6. E. Wang and R. Chow, "What Can I Do Here? IoT Service Discovery in Smart Cities," *Proc. IEEE Int'l Conf. Pervasive Computing and Communication Workshops* (PerCom 16), 2016, pp. 1–6; dx.doi.org/10.1109/PERCOMW.2016.7457097.
7. "Personalized Privacy Assistant Project," Carnegie Mellon University, 2017; www.privacyassistant.org.

Richard Chow is a university research manager and scientist at Intel Corporation. Contact him at richard.chow@intel.com.