**The Privacy Landscape for Pervasive and Ubiquitous Computing**

Jason Hong

**Abstract**

We present a roadmap for privacy research in ubiquitous and pervasive computing for the next decade, highlighting some key challenges and opportunities for addressing privacy concerns. We organize our work using two different lenses. The first looks at privacy from the perspective of devices, characterizing privacy based on different tiers of devices. The second examines privacy from the perspective of different entities involved with privacy, mapping out an ecosystem for privacy. We advance the argument that, today, too much of the burden of privacy is on end-users, and to achieve a sustainable ecosystem we need to develop new ways of shifting this burden to other entities, such as developers, service providers, governments, and third parties.

**Keywords**

Privacy, privacy ecosystem, systems, user interfaces

**Author Bio**

Jason Hong is an associate professor at Carnegie Mellon University focused on usable privacy and security as well as mobile and ubiquitous computing. He is a member of ACM and IEEE.

**Email address**

jasonh@cs.cmu.edu

**Overview**

It's been over 25 years since Mark Weiser first introduced us to ubiquitous computing. If we could time travel and bring someone from 1991 to the present day, they would agree that Weiser's vision of computation, communication, sensing, and actuation being woven into the physical world has happened. A person can now go into any big box store and purchase smartphones, tablets, wearable fitness trackers, webcams, drones, smart thermostats, network-enabled toys, and more.

However, in their initial deployments of ubicomp systems, Weiser and his team also hit upon another issue that we are still struggling with today, namely privacy. There is a fundamental tension with all of this rich sensor and log data that ubicomp devices can collect. On the one hand, this data can be used for personal and societal benefit, in terms of improving healthcare, sustainability, transportation, education, urban planning, finance, and more. On the other hand, these same technologies introduce many new privacy risks, often at a rate faster than legal mechanisms and social norms can adapt.

In a ubiquitously connected world, the costs of collecting, storing, inferring, searching, and sharing data are dramatically lowered. The privacy risks range from everyday ones—such as monitoring by overprotective parents, undesired social obligations with friends and family members, and overly intrusive marketing—to extreme ones, such as threats to civil liberties by governments as well as dangers to one’s personal safety by stalkers, muggers, and domestic abusers. Every day, there is some new headline, interview, op-ed piece, blog post, research paper, or book describing people’s concerns with pervasive computing technologies, regarding the strong potential for abuse, general unease over a potential lack of control, and overall desire for privacy. In some cases, these concerns have even led to outright rejection of systems. For example, Pew Internet found that 54% of app users chose not to install a smartphone app when they discovered how much personal information the app requested, and 30% of app users uninstalled an app after learning that an app was collecting personal information that they did not want to share [3]. In short, privacy may be the greatest barrier to creating a ubiquitously connected world.

In 2001, Mahadev Satyanarayanan outlined a broad research agenda for pervasive computing [24]. Here, in this paper, we take a deep dive into one of the themes he identified, namely privacy. More specifically, we present a roadmap for privacy research in ubiquitous and pervasive computing for the next decade. This roadmap is not comprehensive, as privacy is too broad of a sociotechnical issue for any single article to address. Instead, we highlight some challenges and opportunities that offer good potential leverage in addressing privacy concerns. We organize our work using two different lenses. The first is a pyramid of pervasive computing that describes different tiers of devices. Using this pyramid, we sketch out different challenges for privacy at opposite sides of the spectrum. The second is an ecosystem for privacy that maps out the different entities that are involved with privacy, including end-users, developers, service providers, and third parties. We advance the argument that, today, too much of the burden of privacy is on end-users, and to achieve a sustainable ecosystem we need to develop new ways of shifting this burden to other entities.

**Different Privacy Issues at Different Tiers of Pervasive Computing Devices**

Pervasive computing and its current industry name of Internet of Things is often talked about as a single monolithic concept. However, it’s more useful to think of it as a three-tier pyramid. Each tier represents a different class of device, based on the capabilities of the device as well as our relationship to the device. Each tier also poses different kinds of privacy challenges due to the capabilities of the devices in that tier.

At the top of the pyramid are devices with a great deal computational heft, rich sensing capabilities, fast networking, long battery life, and high interactivity. These devices will be highly personal and be what people typically think of as computers. Example devices here include laptops, smart glasses, tablets, smartphones, and gaming devices. Each person will only have a few of these devices but will also spend a lot of time with them. Most of these devices will have common operating systems, can run third-party software, and will be manufactured by large corporations with a great deal of experience in developing secure software.

In the middle are devices that offer basic interactivity, such as TVs, smart watches, refrigerators, thermostats, electronic whiteboards, cable boxes, and interactive toys. Some of these devices will have advanced sensing and computing capabilities, but the key characteristic here is that people will only use these devices at most a few times a day, and they will also only require a little bit of their attention to use. There will also be greater diversity here in terms of manufacturers, operating systems, and software development experience.

At the bottom of the pyramid there will be hundreds of devices per person, each of which lie far in the background of our attention. These might include RFID-enabled ID cards and badges, clothes, HVAC, digital lightbulbs, smart toilets, smart meters, security systems, implanted medical devices, digital picture frames, cheap environmental sensors, electronic locks, and more. Most of these devices will be embedded or situated in homes, buildings, and public places. Devices in this tier will have very little computational resources, basic sensing, few (if any) software capabilities, and a wide range of software and operating systems. Many of the manufacturers of devices in this tier will also have little experience in developing reliable software and pushing out updates.

|  |  |  |
| --- | --- | --- |
| **Tier** | **Example Devices** | **Characteristics** |
| Top tier | Tablets, glasses, laptops, smartphones | Few devices per personPrimary privacy issue is appsHigh awareness of devices but not of app behaviorsStores highly sensitive data |
| Middle tier | TVs, refrigerators, thermostats, smart toys | Tens of devices per personPrivacy issues are hybrid of the other tiers |
| Bottom tier | HVAC, RFIDs, implanted medical devices, smart toilets, digital picture frames | Combined, hundreds of devices per personPrimary privacy issue is scaleLow awareness of devicesPoor I/O capabilitiesFew capabilities for running privacy softwareLow manufacturer support |

**Table 1.** Three different tiers of pervasive computing.

Here, we sketch out six different dimensions of privacy for describing these tiers. *Awareness* examines how easy it is for people to be aware of both what might be sensed in a given situation as well as what is actually being sensed. *Depth of sensing* means the richness of how much can be sensed by a given device. *Temporal scale* refers to the scope of time in which a device will be sensing data about a person. *I/O* describes how much input and output capabilities a device has for interacting with user interfaces related to privacy, for example configuring policies and viewing log data. *Privacy software* describes how much privacy and security-related software can be run on the device itself. In contrast, *third-party software* describes whether software developed by others can be installed on the device and run. *Manufacturer support* relates to how much ongoing support manufacturers give in terms of fixing bugs and upgrading software.

Each of the three tiers has similar kinds of privacy challenges, but differ primarily in terms of degree and scale. We will focus only on the top tier and the bottom tier, as they represent opposite sides of the spectrum. The middle tier will be a hybrid of these two sides.

For devices in the top tier, people will have high awareness of these devices, primarily because most of them will have stylish form factors and are often readily available for interaction. These devices will have rich sensing capabilities, being able to sense our physical and social contexts. For temporal scale, we will spend a great deal of time with these devices in our daily lives, as we will often carry them as we move from place to place. In terms of I/O, these devices will be able to do text, pointing, speech, and more, and will often have high-end displays that can be used for displaying privacy-related information. For privacy software, these devices will have enough computational power to run necessary privacy-enhancing technologies and security protocols. End-users will also be able to install third-party software on these devices, most likely through app stores. The community’s experience with smartphones suggests that these apps may be the most difficult aspect of privacy for this class of device. Perhaps most importantly, the manufacturers of these devices will have a lot of experience in developing reliable software and in pushing out software updates to fix potential privacy and security software bugs.

In sharp contrast, devices in the bottom tier will be almost the exact opposite. For awareness, devices will be hard to see and often part of the physical infrastructure. For sensing, these devices will typically have few sensors and will be severely constrained by smaller batteries, limiting how much they can sense. In terms of temporal scale, most of these devices will be embedded in specific places, meaning that they will be able to capture a great deal of data over time, though they may not always be able to differentiate between specific individuals. With respect to I/O, these devices will have little or no capabilities, making it very hard to configure and manage privacy. These devices will also have little support for third-party software, typically only running the software that the manufacturer installed on them. Lastly, there will likely be hundreds of manufacturers of these low-end devices, most of which will have little experience in developing software or in pushing out software updates to fix bugs. A likely scenario is that there will be many devices that never get software updates, perhaps due to developer cost, lack of awareness by end-users, or the manufacturer has gone out of business [27]. Perhaps most interestingly, all of these issues will be exacerbated by scale. That is, each of these issues might be manageable with a handful of devices, but quickly become overwhelming when managing hundreds of devices that all have different software, support different networking protocols, and offer different user interfaces.

Given this context, there are a wide range of research challenges for privacy. One of the most salient issues is improving end-users’ awareness of what data is being gathered, where it is being sent (e.g. to Facebook or to a cloud service), and how it is being used. For top-tier devices, people will often already be aware that the device is there. As such, the core issue is awareness of what apps are doing. There is a growing body of research in this area, focusing primarily on smartphones. For bottom-tier devices, there is the same concern about apps, but more important is awareness of the devices themselves, since many of them will be small and possibly even part of the physical infrastructure. Here, we pose a research challenge to the community: can we make it so that when a person enters a room, they can reliably identify all of the sensors and data flows within 30 seconds? This challenge leads to many questions. Should end-users be notified through BLE beacons (similar to the PAWS system [15]), through commonly located visual displays (e.g. near the light switches in a room), or LEDs or audio on sensors? What kinds of common interaction languages can be developed to inform people about sensors? Can we develop cost-benefit models that can balance the benefits of awareness with the costs of user attention and annoyance from repeated or uninteresting notifications? A good example of a step forward towards addressing these challenges is the web site <http://signifiers.io>, which is a project the author helped with that explored different kinds of audio, visual, and motion feedback to convey to users that devices are actively sensing users.

Another privacy issue is with input and output. Top-tier devices will be able to support high-quality user interactions with privacy-related UIs, but bottom-tier devices, with their poor I/O capabilities, will not. Exacerbating this issue is scale. The sheer number of pervasive computing devices will make what would ordinarily be trivial tasks into significant challenges. Configuring privacy for a single device is tractable. Configuring privacy for hundreds of devices, each of which has a different user interface and different sensing capabilities, is not. One research thrust here is developing better tools to help configure and manage multiple devices and apps with respect to privacy. Can we create new kinds of network protocols as well as system architectures that make it easy to add, configure, and maintain devices, all in the context of privacy? Can we also help people make good decisions, for example avoiding misconfigurations as well as making privacy choices that they are comfortable with? Are there ways we can use crowdsourcing or artificial intelligence techniques to help people make good trust decisions as well as quickly surface relevant privacy-related events and potential abuses?

A third privacy issue is that the vast majority of low-end devices will have minimal capabilities in running privacy and security software. Given scalability issues, it makes sense that the infrastructure should offer basic privacy and security functionality. What kinds of services should there be to facilitate the development and deployment of these bottom-tier devices? For example, some essential features might include access control, network security, and software updates. Are there also new ways of conceptualizing these features in the context of pervasive computing? For example, can we use proximity to a person or a room to simplify access control (e.g. “make it so people can always access basic services and sensors in the current room they are in”)? As another example, past work has shown that who is requesting data is an important factor in sharing decisions [17,25]. Although it seems like a privacy paradox, can we use big data approaches to model people and places so as to improve privacy? For instance, can we use sensor data to infer people’s relationships with each other, so as to offer useful defaults for sharing?

**Fostering An Ecosystem for Privacy**

The previous section examined privacy from the perspective of devices. Here, we examine privacy from the perspective of different players that will be involved in pervasive computing deployments.

Today, the burden of privacy is shouldered primarily by end-users. Individuals have to know a great deal about the capabilities and behaviors of their devices, the operating systems and apps they are using, hidden metadata like geotags in photos, where the configuration user interfaces are located and what the right settings are, on top of all of the devices owned by others or embedded in the physical infrastructure that might be monitoring them. In short, individuals must have a great deal of awareness and knowledge as to how to protect themselves, and have to take many affirmative steps to do so

Instead, we argue that we need to expand privacy research to foster an *ecosystem for privacy*. Using an analogy with spam email, in the early 2000s, people had to spend a great deal of time manually deleting spam from their inbox. However, over time, email service providers started to deploy spam filters, Internet Service Providers started to coordinate in developing blacklists to filter out certain IP addresses, and law enforcement worked with industry to take down botnets and arrest individuals that were most responsible for spam. While spam is still an ongoing problem, end-users manually handle far fewer spam emails than they would have to because of the rest of the ecosystem.

Similarly, can we foster an ecosystem for privacy for pervasive computing, one that can help shift the burden of privacy from being solely on end-users? How can we as researchers empower other entities in helping manage and protect privacy for end-users? Examining the ecosystem for smartphones is instructive here, as smartphones are the most developed form of pervasive computing technologies today. Today’s smartphone ecosystem can be broken down into the following entities:

* *Developers*, the people who design and create apps
* *Third-party developers*, the people who create reusable libraries that facilitate other developers
* *Service providers*, the organizations that deploy apps or offer supporting networked services (e.g. advertising, analytics, maps, social networking, cloud storage)
* *App stores*, the organization that aggregate and distribute apps (today, this would be primarily Google, Apple, Amazon, and Microsoft)
* *OS providers* (today, primarily Google and Apple)
* *Hardware manufacturers*, for example Samsung, LG, Motorola, and Apple
* *Government agencies and third parties interested in privacy*, including the US Federal Trade Commission and equivalents, Electronic Frontier Foundation, Center for Democracy and Technology, journalists, and independent web sites like PrivacyGrade.org that rate the privacy of apps

It is likely we will see a similar set of players for pervasive computing. Using this framing, we sketch out some opportunities for privacy research in Table 1. Note that we have combined third-party developers and service providers together, since these two are often the same entity. We also skip over hardware manufacturers in our discussion, in part because we discussed privacy and devices in the previous section. However, from a business perspective, hardware manufacturers are an interesting potential point of leverage for privacy, because they often already have a clear business model in selling the devices themselves rather than an advertising model which entails collecting data about their users. As such, they may be interested in including privacy enhancing technologies with their devices, especially if those technologies can also improve more obvious features like battery life or the user experience.

|  |  |
| --- | --- |
| **Entity** | **Opportunities for Privacy Research** |
| Developers | Checkable purposes for sensitive data useBetter reusable componentsWeaving and enforcing privacy properties throughout an app |
| Third-party developers and Service Providers | Clearer explanations for devs of what their software libraries / services do and whyNew ways of building and deploying privacy-sensitive services New ways of deploying advertisements while minimizing collection of PII |
| App Stores | More effective methods of scanning for potential privacy problems at scaleSimpler summaries of potential privacy issues with apps |
| OS and Middleware Manufacturers | Support for software updates on low-end devices Formal or inferred models of “normal” device behaviorDifferent levels of programming abstractions for accessing sensitive data |
| Hardware Manufacturers | Visual, audio, and physical feedback about sensed dataAligning privacy with power consumption and user experience |
| Government and Third Parties | Tools for checking privacy-related behaviors of appsTools for monitoring behaviors of service providersEasily understandable summaries of apps to raise awareness |
| End-Users | Better understanding of end-user concerns about privacy Simpler mechanisms for awareness and controlBetter ways of expressing preferences |

Table 2. Different entities in the pervasive computing ecosystem for privacy and opportunities for privacy research.

**Developers**

There have been many system architectures and data processing techniques developed for privacy-sensitive pervasive computing systems. However, most of this work has focused on one data type, namely location.

Furthermore, past research has found that app developers have little knowledge of laws or best practices surrounding data collection (e.g. [2]). In short, developers still need a lot more help in creating privacy-sensitive systems. Below, we sketch three promising paths for assisting developers.

Past work has found that people care a lot about who is requesting personal data and why [4,17,18], but there is currently little support for these kinds of explanations. Long privacy policies are burdensome. iOS offers rudimentary support for explaining why sensitive data is being requested, but these explanations are just plain text. What if developers had to specify purposes from a known set? A major challenge would be to develop an agreed-upon ontology, but if we could do that, these purposes could help in automatically generating displays of why an app uses one’s information (e.g. “this app uses your location data for ads” or “this app uses your contact list for social networking”), facilitate the development and use of various tools to check those purposes against the app’s actual behaviors, and enable new kinds of context-based access control (e.g. “allow microphone only for social media and health apps”). These purposes might be supported through new techniques for annotating source code or augmenting manifest files.

Developers also need a high degree of skill in managing all of the sensitive data they collect. Can we create better reusable components that offer useful privacy properties? To use an analogy, databases offer ACID properties: Atomic, Consistent, Isolated, and Durable. A developer doesn’t have to know anything about these properties or anything about the internals of how a database works to gain significant benefits from using one. As such, one research direction would be enumerating a set of useful and desirable privacy properties, as well as creating tools that either have or can support those properties. Another possible starting point might be to understand what kinds of difficulties developers are facing when they are trying to implement privacy features, and look for opportunities for reuse or generalizability.

A related path would be better ways of expressing desired properties or policies, and then weaving those into an application directly or having them be enforced system-wide. Examples for privacy might include direct support for privacy in programming languages (e.g. taint tracking or other kinds of metadata tags on variables), or making it easier to ensure that a set of policies is enforced as the code is compiled [26].

A key challenge for these lines of research is coming up with a list of desired properties that are meaningful and feasible. For example, in security, researchers often talk of confidentiality, integrity, and availability. At a high level, some desired privacy properties might include anonymity, minimal interruptions, avoiding spam, control and feedback about data flows, avoiding behavioral advertising, and preventing embarrassing content from being seen by friends and family. The Fair Information Practices might also offer some other desired properties, such as notice, consent, recourse, and so on. However, it is currently not clear how many of these properties can be implemented. There are promising advances in some of these properties, in particular differential privacy [7], a principled way of adding noise to sets of data so that analysts can still get useful aggregate results without learning information about specific individuals. However, much like Asimov’s Three Laws of Robotics, there is still a very wide technical gap between these high-level declarations and the code needed to implement them.

**Third-party Developers and Service Providers**

If pervasive computing follows the evolution of smartphone apps, we will see a host of third-parties offering useful libraries and services. However, it will also likely be difficult to understand the privacy-related behaviors of these components. For example, past work has found that in some cases, it is not the developers themselves but rather third-party libraries included in an app that are collecting sensitive data [1,2], often to the surprise of developers. From an analysis of smartphone apps on PrivacyGrade.org, we found that about half of the apps that use location data only do so because a third-party library uses it, rather than the app itself.

From the app developer’s perspective, it is very difficult to understand the privacy-related behaviors of these third-party libraries. Today, there is no easy way for a developer to understand what data a library uses and where it sends that data to, beyond reading a lot of documentation. Furthermore, there is no easy way to ensure overall properties of a distributed system. For example, an app might run on a smartphone, be managed by an operating system, use different hardware and software components, and send data to cloud services. How can privacy properties be guaranteed for parts or even all of this system? This is an extremely difficult and open issue for cybersecurity and software engineering in general. To some extent, Android’s approach to having manifest files is one step forward. App developers have to state up front what sensitive data their app might access, in the form of permissions, such as “camera” and “call log”. Can an approach like this be applied to pervasive computing? Are there ways of checking or even proving that an operating system or cloud service is enforcing these permissions correctly as information flows outside of its original device? Can trusted computing bases be used to help?

A related issue is that app developers and third parties need a sustainable way of generating revenue. Advertising-based business models are a popular approach, but pose unique concerns for privacy in that these services have a strong incentive to collect as much data as possible so as to serve up more personalized ads. While there are certain kinds of uses that are clearly unacceptable (e.g. showing job opportunities based on race), it is still unclear what sensitivities people have to different kinds of data and inferences under different situations. We discuss this issue more under the section on end-users. Also, are there better ways of sharing personal data with ad networks and other services so that they can deliver personalized content while also minimizing the sensitivity of personal data that is shared? One good example is Privad [10], which pre-fetched ads so that end-users could see ads, and ad services could still collect clickthrough data but not detailed behavioral information about end-users.

**App Stores**

Given the success of app stores for smartphones, it is very likely we will see similar stores for other pervasive computing platforms. These app stores also present a unique point of leverage for privacy as well, as they may facilitate privacy at scale.

For example, one research opportunity is to develop new techniques for examining the privacy-related behaviors of apps. A number of research projects have looked at static analysis [23], dynamic analysis [8], and even crowd analysis. With respect to crowd analysis, past work has used paid crowd workers to inspect screenshots and descriptions of apps [18,19] or get feedback from actual users of apps [1]. One drawback of dynamic and crowd analysis techniques is scalability, in that they take more time, money, or human attention to work than static analysis approaches. As such, one research direction is to develop new ways of combining these approaches. A likely approach is to use static analysis on all apps, and then using more costly dynamic and crowd analysis for the most popular apps or selected apps flagged by static analysis.

Another research opportunity for app stores is to develop better ways of conveying privacy information to consumers. Today, the state of the art is to offer a long privacy policy full of legalese that few people (aside from lawyers) actually read. Computer-readable formats (such as P3P for web sites [5]) and visualizations (such as privacy nutrition labels [13]) are possible alternatives, but have seen little adoption in practice. A challenge is that app stores want to balance between protecting end-users while also getting people to download more apps, and offering more privacy information may lead to fewer downloads. As such, this kind of privacy information may be better for third parties to offer, though there would also be many challenges here in offering up-to-date information.

**OS and Middleware Manufacturers**

With respect to operating systems, there is the OS for individual devices, and middleware systems for managing multiple devices for homes or entire buildings. There has been a great deal of work looking at privacy issues for the former issue of operating systems, in particular for smartphones (for example, [8,12]). As such, we will focus more on the latter issue.

A core issue here is helping people manage privacy at scale. As mentioned earlier, many low-end devices will have few if any software updates. While this issue is more related to cybersecurity than privacy, poor security can lead to accidental leaks of sensitive information. What kinds of mechanisms can we build so that our middleware can help keep software on low-end devices up to date? Complementarily, can we build formal models or infer models of normal vs abnormal behavior?

One possibility is to consider how to do a division of labor between the middleware and individual devices. For example, one option is to have low-end devices specify a great deal of metadata that the middleware can use as hints. One simple example is to have a URL that points to software upgrades, making it easy to check all devices to see if they have the latest versions. Contrast that today’s process of searching for the manufacturer’s web page, searching for the product ID, downloading the software onto one’s PC, and then installing the updates onto a device. Another simple example is to have a description of what kinds of sensors the device has, what network services it connects to, and for what purposes. An early example of models for device behaviors are Manufacturer Usage Description [16], a draft IETF specification for letting manufacturers define normal behaviors. This kind of metadata can help in generating appropriate notifications, and can help the middleware understand if the device is operating within normal parameters. What other kinds of metadata might be useful to help middleware with managing privacy?

Another core issue for middleware is helping developers. For example, can we offer APIs that make it easier for developers to balance between privacy and utility? As one example, in many situations, it is likely that people will be more willing to share that they are at “home” or “work” (just the string label) versus their exact GPS location. As another example, it is likely most people will be more comfortable with a pervasive computing app asking how loud it is in a given room rather than accessing the raw audio stream. These kinds of programming abstractions offer clear benefit for developers, in that they don’t have to know anything about machine learning, while still letting them get the main information that they are interested in. These abstractions might synthesize or summarize data by space, by time, or by granularity, and it might also be possible to incorporate some model of end-user preferences to help determine sensitivity of the queried data.

**Government and Third Parties**

One underexplored issue for privacy in the context of pervasive computing is in empowering government agencies and interested third parties, such as journalists or advocacy groups, to help take the lead in pinpointing privacy problems and offering meaningful alternatives. With respect to government, one possible area of research is privacy for children. Many countries have laws that govern what kinds of data can be collected about children. Unlike other aspects of privacy, privacy for children is very clear cut, and might also offer leverage points for advancing privacy in general. An example research issue here is developing scalable methods for identifying apps for kids and determining if those apps have inappropriate tracking behaviors [20]. Can we develop similar kinds of analysis tools for devices, in particular toys, to understand their behaviors and to ensure that they are complying with existing laws?

More generally, there are two kinds of research that could help governments in either regulating or fining privacy violations. The first is in developing better kinds of analysis tools that can help government agencies understand the specifics of an app’s or device’s privacy-related behaviors. Much of this kind of analysis today is done manually and often by lawyers rather than computer scientists, and so tools that can help novices can greatly advance things. The second is in measuring uses of personal data to ensure that devices and network services are behaving appropriately. An example might be continuously evaluating advertising networks to understand what data they collect from pervasive computing devices and how they use that data, e.g. not discriminating based on race or that they are complying with their stated privacy policies. Given that pervasive computing devices will likely all be running different software, the only scalable approach is to monitor network traffic, which leads to a host of research questions, such as how to identify specific devices if there are many devices, how to spot potentially sensitive data in network traffic, and how to reliably determine cause and effect, all in a scalable manner. As a concrete example, is there a general tool that could help analysts quickly understand that Samsung’s Smart TV [21] was monitoring voice data, or what exactly Mattel’s Talking Barbie [14] does when you talk to it?

Interestingly, the same kinds of research mentioned above—better analysis tools and better measurements—are just as useful for third parties interested in privacy. In particular, journalists are a particularly interesting case, as they are often interested in exposing privacy surprises, and also offer a unique angle for potentially improving privacy, in terms of being able to shame egregious services or devices.

**End-Users**

While there have been many user studies on privacy in the context of pervasive computing, there is still much that we don’t know. For example, in 2012, the New York Times published an article that described how the big box store Target guessed if a young teenage girl was pregnant, using this data to deliver printed ads offering discounts for baby-related products [6]. This article led to negative comments from people worried about privacy and the growing amount of data that companies have about us. However, from a research perspective, we currently have little insight as to why people felt so concerned. Is it the fact that Target collected this data? That they used the data to infer pregnancy? That they sent coupons for pregnancy-related products? Or that the teenager’s father found out that she was pregnant through these coupons? Currently, all of these issues fall under the broad umbrella of privacy, and unpacking them could help with appropriate designs, interventions, or policy decisions to address people’s concerns.

Similarly, for pervasive computing systems, we need better qualitative and quantitative data to understand the exact nature of people’s concerns, so that we can legitimately address those concerns and draw clear lines about what is and is not acceptable. A deeper understanding of these issues would greatly improve our ability to design pervasive computing systems. For example, how concerned are people in general about different kinds of data types, such as location data, video streams, or sleep data, as well as combinations of data? Following up, how do people’s privacy calculus change based on the different perceived purposes of data use? For example, in past studies, Lin et al [18] found that people were very concerned about contact list data being used for advertising, but mostly neutral about location data being used for social media.

If we can develop such a privacy calculus, can we also use it to develop useful defaults? Defaults are an important but often overlooked issue for privacy. For example, if all sensed data is private by default, the pervasive computing deployment will likely fail because of underutilization, low perceived value, and high burden in terms of end-users having to constantly allow apps to access personal data. In contrast, if all sensed data is publicly visible, there will likely be protests and potentially even outright rejection of systems due to legitimate privacy concerns. If we can develop good defaults for what data should be shared with whom and when, it can likely reduce privacy concerns and reduce end-user burden in terms of configuration, while offering app developers and users of the deployed pervasive computing systems a basic level of utility.

Does gender play a role in people’s perceptions of privacy? For example, are women more likely to share data if it benefits groups or society in general? Also, are women more likely to be worried about tracking, in part due to concerns about stalking? Similarly, do people from different cultures have different conceptions and concerns about privacy? Thus far, the vast majority of research systems and studies have been conducted on WEIRD people [11], that is Western, Educated, Industrialized, Rich, and Democratic. Few studies have examined these kinds of cultural issues for privacy.

How do people’s concerns about privacy change over time? When landline telephone systems were first deployed, many people objected to having landline phones in their homes [9], because it “permitted intrusion... by solicitors, purveyors of inferior music, eavesdropping operators, and even wire-transmitted germs.” Are there better ways to predict how people’s attitudes and behaviors might change?

Lastly, policy makers have also been increasingly adopting the notion of contextual integrity [22] as a working definition of privacy, which looks at appropriate flows of information for specific contexts of use based on norms and values. For example, it makes sense to share personal health data with doctors, but not necessarily with co-workers. However, one major challenge is in aligning often-rigid pervasive computing systems with fluid notions of context and norms. Are there scalable ways of using sensor and log data to help operationalize contextual integrity?

**Conclusion**

In this paper, we have sketched out a research roadmap for privacy for pervasive computing, using two different lenses, with one focused on devices and the other focused on entities in the ecosystem for privacy. We highlighted some of the major privacy challenges as well as research opportunities that have good potential leverage, though it is still an open question as to what are the best points of leverage for achieving privacy in practice. Furthermore, we would be remiss if we did not point out that there are many other privacy challenges in specific domains of pervasive computing, such as lifelogging, drones, social media, and more.

We are currently at a crossroads, not just in computing, but in history. There is only one point in time when the foundation is laid for how computation, communication, sensing, and actuation will be woven into our physical world, and that time is now. These pervasive computing technologies offer tremendous opportunities in terms of healthcare, safety, sustainability, education, and more. But this vision is possible only if we can find ways of addressing the privacy issues, if we can foster trust that the systems we build can respect people as individuals, if we can offer people tangible value, if we give people the right levels of control and feedback, and if these systems do what people expect them to do. We leave you with a parting thought: How can we create a connected world that we would all want to live in?

**References**

1. Agarwal, Y. and Hall, M. ProtectMyPrivacy: Detecting and Mitigating Privacy Leaks on iOS Devices Using Crowdsourcing. *Proceedings of the 11th International Conference on Mobile Systems, Applications and Services (MobiSys ’13)*, (2013).

2. Balebako, R., Marsh, A., Lin, J., Hong, J.I., and Cranor, L.F. The Privacy and Security Behaviors of Smartphone App Developers. *Workshop on Usable Security*, (2014).

3. Boyles, J.L., Smith, A., and Madden, M. *Privacy and Data Management on Mobile Devices.* 2012.

4. Consolvo, S., Smith, I.E., Matthews, T., LaMarca, A., Tabert, J., and Powledge, P. Location Disclosure to Social Relations: Why, When, & What People Want to Share. (2005).

5. Cranor, L.F., Langheinrich, M., Marchiori, M., and Reagle, J. The Platform for Privacy Preferences 1.0 (p3p1.0) specification. 2000. http://www.w3.org/TR/P3P/.

6. Duhigg, C. How Companies Learn Your Secrets. *New York Times Magazine*, 2012. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

7. Dwork, C. Differential privacy: A survey of results. *International Conference on Theory and Applications of Models of Computation*, (2008), 1–19.

8. Enck, W., Gilbert, P., Chun, B.-G., et al. Taintdroid: an information-flow tracking system for re- altime privacy monitoring on smartphones. *The 9th USENIX conference on Operating systems design and implementation (OSDI’10)*, USENIX Association (2010), 1–6.

9. Fischer, C. *America Calling.* University of California Press, 1994.

10. Guha, S., Cheng, B., and Francis, P. Privad: practical privacy in online advertising. *USENIX conference on Networked systems design and implementation*, (2011), 169–182.

11. Henrich, J., Heine, S.J., and Norenzayan, A. The weirdest people in the world? *Behavioral and brain sciences 33*, 2-3 (2010).

12. Jing, Y., Ahn, G.-J., Zhao, Z., and Hu, H. RiskMon: Continuous and Automated Risk Assessment of Mobile Applications. *Proceedings of the 4th ACM conference on Data and application security and privacy - CODASPY ’14*, ACM Press (2014), 99–110.

13. Kelley, P.G., Bresee, J., Cranor, L.F., and Reeder, R.W. A “nutrition label” for privacy. *Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS ’09*, ACM Press (2009).

14. Kim, S. Here’s What It's Like Playing With the Talking ‘Hello Barbie.’ 2015. http://abcnews.go.com/Business/playing-talking-barbie/story?id=33806499.

15. Langheinrich, M. A Privacy Awareness System for Ubiquitous Computing Environments. (2002), 237–245.

16. Lear, E., Droms, R., and Romascanu, D. *Manufacturer Usage Description Specification.* 2016.

17. Lederer, S., Mankoff, J.C., and Dey, A.K. Who Wants to Know What When? Privacy Preference Determinants in Ubiquitous Computing. (2003), 724–725.

18. Lin, J., Amini, S., Hong, J.I., Sadeh, N., Lindqvist, J., and Zhang, J. Expectation and Purpose: Understanding Users’ Mental Models of Mobile App Privacy through Crowdsourcing. *Ubicomp 2012*, (2012).

19. Lin, J., Liu, B., Sadeh, N., and Hong, J.I. Modeling Users’ Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. *Symposium On Usable Privacy and Security (SOUPS)*, (2014).

20. Liu, M., Wang, H., Guo, Y., and Hong, J.I. Identifying and Analyzing the Privacy of Apps for Kids. *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications*, (2016).

21. Matyszczyk, C. Samsung’s warning: Our Smart TVs record your living room chatter. *CNet News*. https://www.cnet.com/news/samsungs-warning-our-smart-tvs-record-your-living-room-chatter/.

22. Nissenbaum, H. Privacy as contextual integrity. *Washington Law Review 119*, (2004).

23. Pandita, R., Xiao, X., Yang, W., Enck, W., and Xie, T. WHYPER: towards automating risk assessment of mobile applications. *Proceedings of the 22nd USENIX Security Symposium, Washington DC, USA*, (2013).

24. Satyanarayanan, M. Pervasive computing: Vision and challenges. *IEEE Personal communications2 8*, 4 (2001).

25. Wiese, J., Kelley, P., Cranor, L.F., Dabbish, L., Hong, J.I., and Zimmerman, J. Are you close with me? Are you nearby? Investigating social groups, closeness, and willingness to share. *Ubicomp 2011 (submitted)*, (2011).

26. Yang, J., Yessenov, K., and Solar-Lezama, A. A language for automatically enforcing privacy policies. *ACM SIGPLAN Notices 47*, 1 (2012), 85.

27. Yu, T., Sekar, V., Seshan, S., Agarwal, Y., and Xu, C. Handling a trillion (unfixable) flaws on a billion devices: Rethinking network security for the Internet-of-Things. *HotNets-XIV Proceedings of the 14th ACM Workshop on Hot Topics in Networks*, (2015).

**Response to Reviewers**

Thank you very much for the insightful comments. I tried to address comments where possible given space limitations or where I had good responses.

**R1 – “In your app store discussion on page 6, I would offer that your comment "as such, this kind of privacy information may be better for third parties to offer" is an interesting idea but could very well lead to accuracy challenges based on synchronization of SW/HW versions as well as a finger-pointing game if things go poorly.”**

I agree with this comment, and added a small caveat that there would still be a lot of challenges in making sure things are up to date.

**R1 - In the very next section, "OS and middleware manufacturers", you mentioned that a great deal of work has already been done around privacy issues for intelligent devices. Perhaps you could toss in a reference to that research.**

I added a reference to TaintDroid and RiskMon as two examples.

**R2 – Arguably, these are in some cases two faces of the same device. Nor are there ever "hundreds" of HVACs per person**

Yes, I can see how someone could interpret this differently, I meant more that the combined number of devices in the bottom tier would be in the hundreds per person, rather than hundreds of HVAC per person. I tried to make this clearer in the table.

**R2 – I also wonder if the bottom tier drawback of "poor I/O capabilities" is all that important if we have smartphones to provide the UIs for us?**

This is a good point, though so far the UIs have been all over the place for today’s devices. Good food for thought, though I don’t have a good or easy way of addressing this question in the paper. I think the reviewer would be ok not directly addressing this issue.

**R2 - BTW, I still receive a lot of spam - these arrive in waves, as it takes our spam filters often a few days to pick up the new patterns and filter new messages out. Hence I am not sure if I would agree with respect to the spam analogy. Maybe this can be somewhat qualified?**

Good point, I tried to qualify this in the writing.

**R2 - Several commments re Table 2: I did not understand what "Checkable" means? Or what the "clearer explanations" signify. Clear for who? Also the term "weaving and enforcing privacy properties" could use some explanation/clarification. The idea that hardware manufacturers provide "Visual, audio and physical feedback" seems a bit naive - one can easily imagine that this won't scale.**

I tried to address as many of these, though the text below the table offers some more details. For example, “checking those purposes” is used, as is “weaving those [policies] into an application directly”, so I would hope readers would be able to make that connection.

**R2 - The idea that one could develop an "agreed-upon ontology" seems too simplistic. This is what P3P tried, and it was a failure, simply because privacy cannot be classified into a few cases.**

I do agree with the reviewer on this. The text originally does allude to the challenges in doing this, but focuses more on what the benefits would be. I think the text, as is, is still ok because it does not claim that it will be easy to do this.

**R2 - I found the discussion around APIs that abstract GPS data from numbers into symbolic places too shallow. "Home" may be just as privacy invasive as GPS coordinates of my home.**

I caveated the text by saying that these abstractions would work for some cases.

**R2 - I found the missing "insights" (section end-users) into the Target case a bit unclear. Why would it be important to know all these things? What would we do differently if we'd know?**

I added some text about this, how we could have better designs, policies, or interventions if we knew how much different parts of the “chain” of concerns affected people.

**R2 - I also found the question of "how concerned are people \_in general\_ about different kinds of data" not convincing.**

I tried to make clearer that the next sentence is a followup sentence. That is, first try to get a general sense of people’s concerns, and then understand how that varies for different situations.

**R2 - Typos, etc:**

Ok I fixed these.

**R3 - But these devices will be most pervasive and I might disagree with the author in that they might not be dumb but powerful and programmable. MAC authentication won’t do the trick of course. Would that turn the pyramid on its head. Would that change the pressure points? If these IOT devices were multi-modal, powerful embedded devices?**

This is a really great comment, though it’s not clear yet what path IoT and pervasive computing will take, and I think the reviewer would be ok with this paper not directly addressing this comment (instead offering it as food for thought).

**R3 - Side note. I liked the idea of signifiers.io (BTW, the author should acknowlegde he is part of that project rather than generally cite it)**

Ok added. I didn’t add myself originally since I had an advisory role and the vast bulk of the intellectual work was by the students, but I see the reviewer’s point.

**R3 - It would have been really interesting if the author could have given some thoughts on what is the likely place that we might be successful and why. What are the necessary incentives to drive this vision forward? Will any of it really pan out? If Yahoo exposed one billion users after 40 years of work then what is the likelihood of one billion sensors being co-opted and exposing the residents of smart city.**

Another great comment, I wish I had solid answers here. I added an extra sentence in the conclusions touching on this issue, that it’s still going to be a long time before we figure things out.