1 Privacy in Ubiquitous Computing Systems as a Social Dilemma

Ubiquitous computing systems that incorporate a wide variety of sensor technologies are an increasingly common part of everyday life for many people. In particular, wearable devices like smart watches and fitness bands, and smartphones carried in a pocket or purse, have been widely adopted. All of these devices include embedded sensors that engage in continuous data collection, and are capable of producing inferences that users consider “extremely private” [18]. For example, in February 2016 a Reddit user posted heart rate data from his wife’s Fitbit activity tracker to enlist the community’s help in troubleshooting what he believed was a malfunctioning device. Instead, he found out from other users that what he had noticed could actually be valid data indicating that his wife might be pregnant (in fact, she was) [10].

Sensor data may seem to users to be non-sensitive and harmless on its own; however, aggregation produces *derived data*: estimates calculated by algorithms that aggregate information collected from multiple users and across time. Derived data consists of new insights and inferences that are not obvious and can be surprising, unsettling or harmful when created and used for purposes users do not expect [17]. Aggregation enables functionality and insights about users that would not be possible from the devices in isolation [2]. Ubiquitous computing systems that involve sensors and data aggregation are socio-technical systems that produce derived data with both positive and negative effects, and the interdependence that produces these effects is as invisible to users as the technologies themselves are.

Because aggregation is necessary for the normal operation of these systems, interdependence between users exists at the technical level by design. But the sensor data is generated and used by people, so the interdependence exists at the social level, too. My activity tracker is able to count my steps more accurately because of the choices of all the other users to wear an activity tracker while walking, jogging, and riding in cars over potholes and bumpy railroad tracks. Therefore, by contributing my data, I am helping to improve how everyone’s activity trackers work.

Prevailing approaches to regulating privacy approach it as an individual right; however, privacy threats are inherently social [19]. Negotiation of boundaries for disclosure must necessarily involve others in some way [16]. We may think about privacy as something that has individual utility, but in reality it serves important social functions. Aggregation and the potential effects of social sorting illustrate that there is such a thing as a collective level of privacy, and there need to be ways to define privacy at the system level so it can be managed collectively [19].

Thinking about privacy as a collective property is especially important in ubiquitous computing systems. For example, the Nest “Learning” Thermostat has a motion sensor; when it stops detecting movement, it switches to Auto-Away mode to conserve energy. However, Nest’s algorithms must identify a threshold for when lack of movement means nobody is at home. The Auto-Away feature has been improved over time by having access to historical data from many households in the Nest cloud. Aggregating this data and identifying patterns across time and households has made the threshold detection more accurate, and allowed all Nest thermostats to decrease the lag time between when the last motion was detected and when Auto-Away mode begins. The sooner it enters Away Mode, the more energy is conserved [11].

Away Mode has positive effects for individual households, in that turning the temperature down in the winter or up in the summer reduces energy costs. However, there are also potentially negative effects. Home/away states can be used to derive other data. Once the Nest learns a household’s occupancy pattern, a number of inferences can be made about the lifestyle and characteristics of the people who live there. A sudden change in activity patterns at night might indicate that someone in the home is having difficulty sleeping, for instance. Combined with data from multiple Nests in the home, a networked camera, or an activity tracker, it might even be possible to determine which member of the household it is, and categorize them as having a higher probability of negative health outcomes due to poor sleep habits. Systems that involve sensors and data aggregation can produce derived data with both positive and potentially negative effects.

Regan [20] found similarities between situations involving information privacy and social dilemmas: “In somewhat the same way that abuses of the environment affect the amount of clean air or water available to others, abuses of privacy by some affect the amount of privacy available commonly.” Pollution of the environment is a social dilemma, because a selfish action by one person (driving to work instead of using public transportation) has negative effects on third parties that are not involved in the decision about how to get to work (using up fossil fuels and creating greenhouse gases). It is in each individual's independent interest to behave selfishly and drive to work because it is faster and more convenient, but if everyone chooses that action all will suffer negative consequences. This is an example of a social dilemma called the tragedy of the commons [7], in which everyone must choose to act against rational self-interest and cooperate to manage a limited resource in order for all to benefit from it. More specifically, it is an example of a common-pool resource problem, which is typically found in situations where people must figure out how to share a natural resource. Common pool resource problems have two characteristics. The first is *excludability*: the ability to prevent others from using the resource. In a common-pool resource problem, excludability is difficult to accomplish. The second characteristic is *subtractability*: the degree to which consumption of the resource by one person affects the availability of the resource for others. Common-pool resources have high subtractability.

2 Derived Data as a Common Pool Resource

It is clear that natural resource systems can present common-pool resource problems. However, social dilemmas exist in information systems as well. Information systems are different from natural systems in that the resource is the information itself, and therefore the resource is produced by humans rather than the natural environment [3]. Also, information is not thought of as subtractable: many people can use the same digital information at the same time without affecting others' ability to use it [22]. This means that online communities like Wikipedia are public goods rather than common pool resources. Wikipedia pages have low subtractability and high excludability, because multiple people can access and benefit from them at the same time. However, there is a cost to the resource: people must invest time and energy creating and editing Wikipedia pages without direct compensation. Angst [1] argues that, like Wikipedia, there are benefits from aggregation and derived data from health records exchanges that constitute a public good. He suggests that while there may be privacy costs to the individual from a national health information exchange, there are also benefits to society. The benefit only manifests, though, if individuals are willing to accept the privacy risks and contribute their data.

Ubiquitous computing systems (like activity trackers) that involve sensors and derived data and depend on aggregation to perform their function exhibit properties of a common-pool resource system. Using the system creates the resource: the derived data. Derived data is non-excludable in that it is difficult for users to prevent others from joining the system and contributing data; one user cannot prevent another from buying and using a Fitbit. It is also subtractable, in that as more users contribute more sensor data, negative privacy-related inferences about users from a larger and more diverse dataset become more likely. This is a difference from how we typically think about subtractability, from consuming or "using up" a resource, to considering the overall benefit and whether it is reduced as the resource is used.

Ostrom [14] presented a framework for analyzing a social-ecological common pool resource system, like pollution of the environment. She wrote that it is important to analyze the resource system to identify its unique structural details in order to solve the dilemma. This includes variables like the characteristics of the natural environment, the resource units, the users of the system, and the current method for coordination and regulation. Derived data in ubiquitous computing is part of a socio-technical resource system, not a social-ecological one. However, I believe this framework of variables applies.

Compare this with an example of a more traditional, yet man-made common pool resource system: a library. An individual user might benefit from checking out as many books at once as possible. However, this prevents others from using the books. In this case, benefit for one user removes the opportunity to benefit from the other users: checking out books is only a benefit to the person who has the books. Unlike library books, derived data in ubiquitous computing systems is both a benefit and a potential harm for all users at the same time. It has positive consequences for other users, in that it is an important aspect of how the system functions and can be used to further improve the system. However, it also puts users at risk for inferences about personal characteristics that could have negative consequences due to social sorting. In other words, each person that uses the system for its positive, intended effects means provisioning data that also makes the negative consequences possible for all users.

In the library example, a mechanism must be devised to regulate the flow of resource items (books) between users so that everyone has a chance to benefit. Typically this looks like time limits for keeping books and a recall procedure for when demand is high. Like the library example, a mechanism could be devised to regulate the derived data, so that users can have the benefits without the harms. Unlike natural resource systems in which characteristics are determined by nature (e.g., in a forestry resource system the growth rate of trees cannot be altered by humans), the technical characteristics of socio-technical systems are designed, and can be changed.

In today's ubiquitous computing systems, choices made by users may be perceived as independent because the technology is invisible. But, they are actually interdependent. A system could be designed to allow its users to coordinate or signal to each other about what kinds of uses of derived data are acceptable, in order to retain the positive benefits that come from offering certain kinds of information for use by the system, and disallowing other uses that violate norms [12]. The physical privacy equivalent might be something similar to looking away while someone types in his password. Knowing that a password is being entered may not be considered a privacy violation, but witnessing the details might be. This represents a shift from personal preference and responsibility for privacy, to group coordination; from thinking about data sharing as an individual choice about what people should or shouldn't disclose about themselves, to a negotiation about what people should or shouldn't know about others.

Others have noted the interdependence of user actions in privacy and security, as well. For example, Regan [20] argued that interdependence in information systems is "making it hard for any one person to have privacy without all persons having a similar minimum level of privacy", and drew connections between privacy and natural common pool resource systems. Dong et al. [4] proposed a system to prevent security breaches in which a group of users might agree to use software to automatically monitor each others' computer behavior and network traffic, and as a result of that visibility, track historical patterns and identify compromised systems automatically. Garg et al. [5] suggested developing monitoring technologies that would automatically inform users when social privacy violations have occurred in online social networks. These two papers advocate creating automated monitoring algorithms; however, a privacy solution like this would suffer from the problem of defining what a violation or breach looks like, just as other automated approaches have. The authors were also not specific about how privacy, as a global property of all information systems, is non-excludable and subtractable.

Ostrom herself wrestled with how to take the lessons she learned from localized natural resource systems and apply them to global common-pool resources, particularly in the area of climate change; this remains an unsolved problem [15]. Ostrom's work demonstrates that every common pool resource system is different, and the details of each system matter for the kinds of management solutions that will work. The approach that this project takes is that it is important to create support for coordination among interdependent users within a ubiquitous computing system, so that the details of managing the derived data as a common pool resource can be specific to the details of the particular situation.

3 Conceptual Structure of the Social Dilemma

Ostrom [14] presented a framework for analyzing a social-ecological common pool resource system. She wrote that it is important to analyze the social-ecological system to identify the specific structural details of the unique resource system in order to solve the dilemma. This includes variables like the characteristics of the natural environment, the resource units, the users of the system, and the current method for coordination and regulation. Derived data in ubiquitous computing is part of a socio-technical resource system, not a social-ecological one. However, I believe this framework of variables applies. Below I describe each part of the framework, how it is represented in the socio-technical system under consideration, and how it will be instantiated in this project.

Resource system. The physical attributes and properties of an ecosystem such as type (water, forest, fish), and its boundaries with other resource systems, size, productivity, and location. For this project, the properties of the resource system are specified as part of the design of the platform for the project, and include the hardware and software components, network connectivity and the data processing and sharing capabilities. The properties of the platform will be based on real systems available today for home automation.

Resource units. The properties of the things that are consumed in the resource system such as their mobility, growth or replacement rate, interaction, economic value, spatial and temporal distribution. For this project, the resource units are the derived data, and the algorithms and processes by which they are produced and used. As with the resource system, the properties of the resource units are specified as part of the problem space of the project.

Governance system. The details of the mechanism for regulating the system, such as the network structure, operational rules, monitoring and sanctioning processes. In a derived data situation, the governance system is who decides how the information is gathered, processed and used; how they decide; and how they get permission. In real world systems, this is determined by system designers and operators and is currently based on notice and choice. The characteristics of the governance system (coordination mechanism) will be addressed as part of the research and design activities.

Users. The characteristics of the people involved in the system, including how many there are, history of use of the resource, where they are located, norms, social capital, how important the resource is for their survival, and their knowledge/mental models having to do with the resource and its use. The users of both social-ecological and socio-technical systems are people, and so all of these attributes are relevant. A particular point of focus for this project is the social norms associated with the creation and use of derived data.

Interactions. The kinds of interactions that can take place in the resource system, including resource use, information sharing, deliberation processes, and conflicts among users. Creating a mechanism for these interactions to take place is part of the design problem. This project will create new, more visible points of interdependence and evaluate their effects.

Outcomes. These are measures of the the performance of the system as a whole and for the individual users: social performance measures, ecological (technical) performance measures, effects on other resource systems. These are the dependent variables for the project, and include things like how effective the coordination mechanism is, measures of perceived privacy, and satisfaction with the system.

4 Norms for Privacy

Privacy is usually discussed in terms of individual preferences for control and disclosure [12], not social norms [8, 16]. Because of the focus on self-management, once private information has been disclosed, “no one could be faulted for using the private information, because it has been made freely available by the owners themselves” [16]. However, in writing about norms for privacy, Nissenbaum [13] said that even in public places there are norms for privacy; for example, it is reasonable to refuse if a stranger asks you your name. People learn rules from others in their family or organizations they belong to for what private information looks like, and how it should be managed [16]. Privacy-related norms, for example, cause people not to stare at others’ body parts under some circumstances [21]. In a study of mobile phone use in public by pairs of people (one of whom is on the phone while the other waits), Humphreys [9] mentioned a social norm against listening in to others’ conversations, although there is “freedom to listen in” if the person at the other end of the call is known to both parties.

It is difficult to imagine what norms for information privacy look like, because information privacy is typically treated as an invisible exchange between individuals and institutions. When a behavior isn’t visible to others, people tend not to make norm-consistent choices [6]. Right now, collected and derived data are not visible to users in ubiquitous computing systems, and so privacy-related norms for how it should be used are irrelevant for user behavior. Users have no choice but to be selfish and norm-free because the system doesn't provide support for coordination and social awareness.

This project has two high-level goals. The first is to discover and validate privacy norms for awareness of information derived about others from sensor data in a ubiquitous computing system, where the derived data is a necessary part of the normal operation of the system. The second goal is to design and evaluate a mechanism that allows a group of users to coordinate around the norms for what information should be known about others, and govern its use. Unlike current approaches to information privacy which focus on individual notice and consent, which are difficult to accomplish in a ubiquitous computing system, this approach will allow users to make decisions about how to manage the common pool resource of derived data for themselves.

The main design question this project addresses is how to represent and measure uses of derived data, so that users can monitor the resource, coordinate about uses of derived data, and maintain accountability. It is difficult to manage a common pool resource in which the resource units are hard to measure [15]; likewise, users of a ubiquitous computing system must be informed some way about the possible uses so they can coordinate about which are acceptable and which are not. Demonstrating that derived data can be collectively managed like a common pool resource will point to new kinds of solutions for digital privacy issues, focusing on the social dilemma aspects of the problem.

References

[1] Corey M Angst. Protect My Privacy or Support the Common-Good? Ethical Questions About Electronic Health Information Exchanges. *Journal of Business Ethics*, 90(S2):169–178, January 2010.

[2] Michael Brown, Tim Coughlan, Glyn Lawson, Murray Goulden, Robert J Houghton, and Richard Mortier. Exploring Interpretations of Data from the Internet of Things in the Home. *Interacting with Computers*, 25(3):204–217, March 2013.

[3] Daniel H Cole. Learning from Lin: Lessons and Cautions from the Natural Commons for the Knowledge Commons. In Brett M Frischmann, Michael J Madison, and Katherine J Strandburg, editors, *Governing Knowledge Commons*, pages 45–68. New York, June 2014.

[4] Zheng Dong, Vaibhav Garg, L Jean Camp, and Apu Kapadia. Pools, clubs and security: designing for a party not a person. In *Proceedings of the 2012 workshop on New security paradigms*, pages 77–86, 2012.

[5] V Garg, S Patil, A Kapadia, and L J Camp. Peer-produced privacy protection. *2013 IEEE International Symposium on Technology and Society (ISTAS)*, pages 147–154, 2013.

[6] Vladas Griskevicius, Joshua M Tybur, and Bram Van den Bergh. Going green to be seen: Status, reputation, and conspicuous conservation. *Journal of Personality and Social Psychology*, 98(3):392–404, 2010.

[7] Garrett Hardin. The tragedy of the commons. *Science*, 162:1243–1248, 1968.

[8] James Harper. Privacy-Invasive Technologies and Their Origins. In William Aspray and Philip Doty, editors, *Privacy in America: Interdisciplinary Perspectives*, pages 113–135. The Scarecrow Press, Inc., Lanham, MD, 2011.

[9] Lee Humphreys. Cellphones in public: social interactions in a wireless era. *New Media & Society*, 7 (6):810–833, 2005.

[10] Amanda Jackson. Husband and Wife Never Expected Their Fitbit Would Tell Them This, February 2016. `http://www.cnn.com/2016/02/10/health/fitbit-reddit-pregnancy-irpt/`.

[11] Farhad Manjoo. The World's Best Thermostat Just Got Better, October 2012. URL http://www.slate.com/articles/technology/technology/2012/10/nest_thermostat_the_ingenious_heating_and_cooling_system_keeps_getting_smarter_.html.

[12] Gary T Marx. Turtles, Firewalls, Scarlet Letters, and Vacuum Cleaners: Rules about Personal Information. In William Aspray and Philip Doty, editors, *Privacy in America: Interdisciplinary Perspectives*, pages 271–294. The Scarecrow Press, Inc., Lanham, MD, 2011.

[13] Helen Nissenbaum. Toward an Approach to Privacy in Public: Challenges of Information Technology. *Ethics & Behavior*, 7(3):207–219, September 1997.

[14] Elinor Ostrom. A diagnostic approach for going beyond panaceas. *Proceedings of the National Academy of Sciences*, 104(39):15181–15187, 2007.

[15] Elinor Ostrom, Joanna Burger, Christopher B Field, Richard B Norgaard, and David Policansky. Revisiting the Commons: Local Lessons, Global Challenges. *Science*, 284(5412):278–282, April 1999.

[16] Sandra Petronio. *Boundaries of Privacy: Dialectics of Disclosure*. State University of New York Press, Albany, NY, 2002.

[17] President's Council of Advisors on Science and Technology. Big data and privacy: a technological perspective. United States Executive Office of the President, May 2014.

[18] Amon Rapp and Federica Cena. Personal Informatics for Everyday Life: How Users Without Prior Self-Tracking Experience Engage with Personal Data. *International Journal of Human-Computer Studies*, 94:1–17, 2016. doi: 10.1016/j.ijhcs.2016.05.006.

[19] Priscilla M Regan. *Legislating Privacy: Technology, Social Values, and Public Policy*. The University of North Carolina Press, Chapel Hill, NC, 1995.

[20] Priscilla M Regan. Privacy as a Common Good in the Digital World. *Information, Communication & Society*, 5(3):382–405, 2002.

[21] Ferdinand Schoeman. Gossip and privacy. In Robert F Goodman and Aaron Ben-Zeev, editors, *Good Gossip*, pages 72–82. University Press of Kansas, Lawrence, KS, 1994.

[22] Gary M Shiffman and Ravi Gupta. Crowdsourcing cyber security: a property rights view of exclusion and theft on the information commons. *International Journal of the Commons*, 7(1):92–112, February 2013.