Proactive Data Sharing to Enhance Privacy in Ubicomp Environments

Dirk Balfanz
balfanz@parc.com

Philippe Golle
pgolle@parc.com

Jessica Staddon
staddon@parc.com

1. INTRODUCTION

Discussions about privacy often assume an antagonistic relationship between those who collect data and those about whom data is collected. Those who collect data (the data consumers) want the greatest possible access to data, whereas those about whom data is collected (the data producers) want privacy to the greatest possible extent.

Much work on privacy has consequently been devoted to designing tools and protocols that allow data producers to safeguard their data from the prying hands of data consumers. More recently, work predicated on some amount of trust between data producers and consumers offers tools for negotiating privacy [10], or recognizes that negotiations of privacy boundaries should not consist only of a defensive posture biased toward minimum disclosure [8].

We propose to heal the relationship between data producers and consumers even further, and take the position that in ubicomp environments, privacy can be facilitated by collaboration. Indeed, we argue that in ubicomp scenarios, the best response to data consumers' desire for information is for data producers to voluntarily provide the data desired.

In a nutshell, the argument goes as follows. The data that is most valuable to data consumers is typically simple, specific, well-defined and not very sensitive from the view-point of privacy. A typical example may be the time you spend inside a store, the brand of coffee you drink or your favorite color for clothes. Unfortunately, it is often not possible to collect this valuable data directly. Instead, data consumers must rely on monitoring equipment (audio, video or other sensors) that produce mountains of raw data from which a few nuggets of valuable data can be extracted, distilled or aggregated. This process is expensive, inefficient and produces as by-product vast amounts of raw data that is useless to many data consumers, yet that may represent a serious threat to the privacy of data producers.

This privacy threat could be avoided simply if data producers were to share data about themselves voluntarily and directly with data consumers. Indeed, voluntary sharing of data would eliminate the incentive for data consumers to set up and operate ubiquitous monitoring systems that gather the same information inefficiently. The privacy of consumers may be enhanced overall, since no extra data would then be collected other than what the data consumers were interested in. In other words, our response to organizations' desire for information is simply to give them the information they want, and only that information, thus preempting the collection of raw data that would be more damaging to privacy.

The following examples illustrate our approach:

- A store may only be interested in learning how much time its customers spend inside the store. There may be no better way to get that information than to set up video cameras at the entrance (and exit) of the store. The video cameras collect a lot more information than the store needs, and that extra information may be open to abuse. A better approach would be for customers to measure themselves the time they spend inside the store, and report that information voluntarily.

- The use of sensors alongside roads has been proposed to monitor congestion and relay information to drivers to improve the flow of traffic. Once installed, these sensors may be used for other purposes, such as for example, catching speeding vehicles. To avoid the possibility of such privacy-invading "mission creep", a better approach may be to never deploy sensors in the first place and rely instead on drivers to voluntarily report congestion.

We return to these examples in section 3. We feel that our approach to privacy is particularly well suited to data collected in a ubicomp environment for two reasons:

1. Ubicomp environments allow for nearly uninterrupted collection of rich data. Once deployed, ubicomp data sensors risk being used for purposes beyond those for which they were originally designed. This creates strong incentives for data producers to share their data willingly, to preempt the deployment of sensors. Contrast this situation with the collection of census data: there is little incentive for citizens to freely offer demographic data to the government (our approach therefore does not apply), but the census survey conducted every few years does not pose nearly the same threat to the privacy of individuals as a ubicomp sensing infrastructure.

2. Valuable information extracted from raw sensor data is usually directly available from data producers. The raw data collected by ubicomp sensors are often distilled into a single valuable fact about one entity: for example, this vehicle travelled from point A to point B, or this shopper likes this brand. This fact is typically known to the entity to whom it pertains, and that entity could volunteer the information directly if asked. Again it helps to contrast this situation with census data: our approach fails there, because there is no single "average citizen" entity that the

census bureau could go to, to ask for mean and median data. Instead there is no way around collecting data for every individual and aggregating it afterward.

Our approach to privacy in ubicomp environments is similar in spirit to the self-regulations that have been successfully put in place by a number of industries (e.g. the advertising industry via the National Advertising Review Board, or the alcohol industry) to preemptively fend off possibly less desirable regulation from the government. These industries find themselves in the same "weak" position with respect to the legislative arm of government as that in which data producers are with respect to large organizations thirsty for data. Self-regulation is a form of preemptive collaboration that helps deflect calls for stricter regulation. In much the same way, we propose to give data consumers the data they want, to preempt the deployment of more privacy-invasive data collection tools.

Organization. In the next section, we review and justify the assumptions that underly our position. In section 3, we offer some more detailed examples of how our position may be applied in practice. In section 4, we discuss limitations of our approach. Finally, we conclude in section 5.

2. ASSUMPTIONS

We define three categories of entities that interact with data:

- **Data consumers:** entities who make use of data collected about other entities. Stores, marketers, regulatory agencies or governmental intelligence agencies are all potential examples of data consumers.
- **Data producers:** entities about whom, or about whose actions, data is collected. Examples of data producers include shoppers, airline passengers, automobiles and more generally any entity that produces a trail of data. Note that we use the terminology of producer and consumer consistently with respect to data. Thus we refer to shoppers (i.e. consumers of goods) as *producers* of data.
- **Data collectors:** entities who collect data from data producers and make it available, in raw or processed form, to the data consumers. In this paper, we focus mostly on ubicomp data collection tools, such as video cameras, arrays of sensors, RFID readers, etc, but data collection may also involve a simple database of transactions.

These categories overlap and a single entity may belong to different categories at different times, or even simultaneously, depending on the roles it assumes. For example, shoppers at a supermarket produce a wealth of data about items they buy, coupons they use, etc. The supermarket may be a consumer of that data (e.g. using it to predict future demand) as well as a data collector with respect to another data consumer if, for example, it makes the data available to its suppliers, or even the police.

We argue that in a ubicomp setting, the overall privacy of data producers would be enhanced by eliminating data collectors to the maximum extent possible, and instead letting data producers share data directly with the data consumers to whom such data is valuable. We justify this position with the following observations. Each observation is followed by a short example to illustrate our point. We offer more detailed examples in the following section.

1. Ubicomp data collection is very inefficient.
The vast amounts of raw data collected by audio, video and other sensors are for the most part of no value to data consumers. The process of extracting valuable information from raw data, through synthesis and aggregation, is as inefficient as looking for the proverbial needle in a stack of hay. The ratio of valuable distilled data (e.g. this shopper likes this brand) to raw data (e.g. hours of videotape footage) is characteristically very low.

2. The graver threat to privacy is raw data.
The distilled, aggregate data that is valuable to data consumers is often far less threatening to the privacy of data producers than raw data. This is particularly true of the "low-yield" raw data collected by ubicomp sensors. For example, anecdotal evidence suggests that most shoppers would rather make public their brand preferences than video footage of them walking around a store (while possibly quarrelling with their partner, disciplining their child, picking their nose, etc.)

3. Data producers are the best source of data.
It is easier, cheaper and more reliable for data consumers to obtain data directly from data producers, rather than rely on data collectors as intermediaries. This assumption holds especially well in ubicomp scenarios, where the costs of setting up and maintaining an infrastructure to collect raw data are typically high. As already noted, there are further costs involved in aggregating and interpreting raw data. By contrast, obtaining valuable data directly from data producers may be no more complicated than just asking for it.

4. Data producers will willingly share data.
In a ubicomp environment, there is a two-fold incentive for data producers to share data about themselves with data consumers. First, sharing some data willingly may be seen as a proactive move to ward off the deployment of more privacy-invasive data sensors that would be used to collect the same data (and more). Secondly, data producers could be enticed to share data with rebates, discounts, or simply a clear explanation of the purpose for which the data is collected and the associated benefits. Thus for example, drivers opposed to cameras monitoring the speed of vehicles on a road may voluntarily contribute information about their speed, to improve the overall flow of traffic and undermine the business case for installing cameras along the road.

We believe that it is in the best interest of both data producers and data consumers to exchange data directly, bypassing data collectors. The privacy of data producers is better protected (by observation 2), while at the same time data consumers obtain more reliable data at cheaper cost (by observations 1, 3 and 4).

Note that the conditions above are not always met and we do not advocate the use of our approach to privacy in every setting. There are scenarios (notably medical data), where privacy must be preserved unconditionally and a voluntary

approach to sharing data is not appropriate (we discuss further limitations of our approach in section 4). However, our approach appears promising in many ubicomp scenarios. We give next some more detailed examples.

3. EXAMPLES

3.1 Traffic Monitoring

Radio-Frequency (RF)-based toll collecting systems such as EZ-Pass (popular on the East Coast) or Fastrak [4] (in the San Francisco Bay area) use *tags* (little boxes subscribers install in their cars) that are read as subscribers drive their cars through toll booths.

The same technology could be used to monitor traffic jams – tag readers could be installed alongside highways to measure the speed of cars and traffic density. While such a system could be used to accurately monitor and report traffic conditions, it would also raise privacy concerns: the same system could conceivably be used to monitor individual drivers' speeds. The existing toll booth installations have in fact raised worries among their users because they were used for more than just toll collection [11].

Our approach suggests a different solution for traffic monitoring. Instead of putting sensors along the road, we propose that cars voluntarily report their speed to a traffic monitoring agency (perhaps through a cell phone that the driver is wearing). To thwart concerns that this could be used to track drivers' speed, we could build the system such that speeds are only reported if they are below a certain threshold, say 80% of the speed limit on a certain road. This way, the system could not be used to track speeders, but it would still accurately report traffic jams.

3.2 Toll Collection

While RF-based toll collection systems, if used for other purposes, can raise privacy concerns (as explained in our previous example), they also serve as a case in point for our proactive data sharing approach. As far as toll collection is concerned, they are an example of a voluntary opt-in system. Subscribers have to get an RF tag and install it in their cars. They understand that whenever they drive through a toll booth, an event is logged in the system.

Compare this with the "congestion charge" system installed in the city center of London [7]. There, cameras monitor every vehicle that enters the city center, taking a snapshot of the license plate. At the end of each day, the system compares the observed vehicles with those for which it has received payment for that day. The owner of each vehicle that has been observed in the city center, but has not been paid for, receives a bill in the mail.

Both (RF-tag and camera-based) systems do the same thing: they provide a way to charge drivers who travel certain routes. In the RF-tag-based system, the drivers essentially generate toll-booth-crossing events voluntarily (by agreeing to install the RF-tag in their cars), and subsequently, the information logged is fairly limited: time of crossing, id of the RF-tag that crossed, and id of toll booth that was crossed. In the camera-based system (in which drivers do not voluntarily share this kind of information), what is logged is a *picture* of the car entering the city center, potentially carrying much more information (for example *who* was in the car).

While there certainly were practical reasons to install a camera-based congestion charge system in London, it seems clear that such a system raises more privacy concerns than an equivalent RF-tag-based solution such as the Fastrak system found on San Francisco bridges. We point out that (the more privacy-preserving) Fastrak uses a proactive data sharing approach, while the London congestion charge system does not.

3.3 Collection of Marketing Data

Another potential use of or approach is to collect marketing data for retailers. Imagine video cameras in the window of a mall store filming people as they pass by[1]. Such information allows the store to evaluate the effectiveness of their window displays for attracting customers, determine shopping time and get some sense of the store's competition (e.g. by noting where departing customers head) but it also potentially includes unnecessary, privacy-compromising information such as a visit to the doctor's office across the breezeway.

With our approach the data producers provide the desired marketing data, possibly in enhanced form, while keeping the unnecessary information private. For example, a data producer can opt to sell their location information while shopping but withhold information during all other activities. This provides far more valuable data to retailers as they can recover complete shopping histories, including the sequence of stores visited and the time spent in each. Amassing this information with video cameras requires collaboration with multiple stores and requires the processing of much extraneous footage.

4. LIMITATIONS

While our argument could be construed to imply that ubicomp is a barrier to privacy, we assert instead that the two can peacefully coexist by means of our ubicomp-*enabled* approach. Mobile data producers can leverage the existence of a ubiquitous computing infrastructure to make their data available in real time to the data consumers. Indeed the true barrier is simply acceptance; the system only works if it is broadly adopted by both consumers and producers.

Economics can facilitate acceptance by the consumers. If producers offer the right data at a better price than the data collectors, there is an incentive for consumers to purchase the data directly from the data producers. In addition, the applications that we envision for this approach involve very simple data queries (e.g. location information) that can be provided directly by the producer with a low error-rate or measured by a producer-enabled tool (e.g. GPS). Hence, the producers can provide reliable data. Finally, there may be public relations benefits to the consumers of adopting this privacy-enhanced approach.

Numerous studies have shown that producers are easily incentivized to provide information [1, 5, 9]. Supermarket "loyalty" cards are just one tool through which producers reveal data in return for rewards (e.g. discounts). The same techniques can be used to encourage the adoption of our approach. For example, stores can provide discounts and early

[1]Although currently video cameras are also frequently used for store surveillance, in the future they may be replaced for that purpose by other technologies such as RFID tags. However, video cameras are likely to always be a candidate tool for market data collection.

sale notice to producers who sell their data. Data consumers are typically interested in producers' data because they provide a service to producers. Hence, they may reward the producers by providing discounted or enhanced services.

To cement acceptance by consumers and producers and generate a stable equilibrium in which neither party deviates from behaving cooperatively, legal measures may need to be coupled with our approach. In particular, since ubicomp advances have made undesirable methods of data collection (see Section 3) easy from a technological standpoint, it may be impossible to ensure that greedy data consumers will not pursue those methods to obtain more data than they would from consumers, other than through legal means. Our collaborative approach to data sharing may in fact facilitate the introduction of legal restrictions on ubicomp data collection without the consent of data consumers. Indeed, since our approach provides consumers with economical access to the desired data, the addition of legal impediments to acquiring the data through other means appears fair.

5. CONCLUSION

We have presented a simple, user-driven approach to privacy in a ubicomp world. Our central tenet is that producers and consumers are not as at odds when it comes to privacy as many believe. Much of the producer data that is valuable to consumers is data the producers are willing to share. In addition, there are numerous advantages to consumers to collecting data directly from the producers including economic, efficiency and public relations benefits. Further, the addition of incentives facilitates adoption by producers.

6. REFERENCES

[1] A. Acquisti. Privacy in Electronic Commerce and the Economics of Immediate Gratification. To appear in *Proc. of ACM Electronic Commerce Conference (EC 04)*, New York, NY, 2004.

[2] J. Canny. Collaborative filtering with privacy. In *IEEE Symposium on Security and Privacy*, pp.45–57, Oakland, CA, May 2002.

[3] J. Canny and Y. Duan. Designing for privacy in ubiquitous computing environments. `http://www.cs.berkeley.edu/~duan/research/drafts/linkana.pdf`

[4] FasTrak Electronic Toll Collection System. `http://www.511.org/fastrak/`

[5] Jupiter Research. Seventy percent of US consumers worry about online privacy but few take protective action, 2002. `http://www.jmm.com/xp/jmm/press/2002/pr_060302.xml`

[6] M. Langheinrich. Privacy by Design – Principles of Privacy-Aware Ubiquitous Systems. In *Proc. of Ubicomp 2001*, pp. 273–291, Springer-Verlag LNCS 2201, 2001.

[7] T. Litman. London Congestion Pricing. `http://www.vtpi.org/london.pdf`

[8] D. J. Phillips. Context, identity, and privacy in ubiquitous computing environments. In *Proc. of Ubicomp 2002 Workshop on Socially-informed Design of Privacy-enhancing Solutions in Ubiquitous Computing,* Goteborg, Sweden, 2002.

[9] S. Spiekermann, J. Grossklags and B. Berendt. E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior. In *Proc. of the 3rd ACM Conference on Electronic Commerce, EC'01*, pp.38–47, 2002.

[10] L. Cranor, M. Langheinrich, M. Marchiori and J. Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. W3C Recommendation, April 16, 2002. `http://www.w3.org/TR/P3P/`

[11] Todd Wallack. They Know Where You've Been – Data collected from FasTrak drivers raise privacy concerns. San Francisco Chronicle, February 12, 2001.