Privacy Tradeoffs: Myth or Reality?
(Panel Summary)

Rebecca N. Wright (moderator) *
L. Jean Camp †
Ian Goldberg ‡
Ronald L. Rivest §
Graham Wood ¶

Abstract. We discuss tradeoffs between privacy and other attributes such as security, usability, and advances in technology. We discuss whether such tradeoffs are inherent, or if it is possible to "have it all."

"You have zero privacy anyway. Get over it."
— *Scott McNealy, 1999*
Chief Executive Officer
Sun Microsystems

Changes in technology are causing an erosion of privacy. Historically, people lived in smaller communities and there was little movement of people from one community to another. People had very little privacy, but social mechanisms helped prevent abuse of information. As transportation and communications technologies developed, people began to live in larger cities and to have increased movement between communities. Many of the social mechanisms of smaller communities were lost, but privacy was gained through anonymity and scale.

Now, advances in computing and communications technology are reducing privacy by making it possible for people and organizations to store and process personal information, but social mechanisms to prevent the misuse of such information have not been replaced. While a major issue in computing and communications technology used to be how to make information public, we now have to work hard to keep it private. The main causes for this are the reduced cost of data storage and the increased ability to process large amounts of data. Nonetheless, one should not entirely abandon the hope for privacy. Rather, new ways of thinking are needed to find solutions based on a combination of technology, policy, and education to try to maintain or increase privacy and to provide new social mechanisms. The problem is

* Stevens Institute of Technology, Hoboken, NJ 07030
† Kennedy School of Government, Harvard University, L 213, 79 JFK Street, Cambridge, MA 02138
‡ Zero-Knowledge Systems, 888 de Maisonneuve East, 6th Floor, Montreal, Quebec, Canada H2L 4S8
§ Massachusetts Institute of Technology, 545 Technology Square, Room 324, Cambridge MA 02139
¶ Appleby Spurling & Kempe, Cedar House, 41 Cedar Avenue, Hamilton, HM 12 Bermuda

Appears in *Proceedings of Financial Cryptography 2002*, Springer LNCS 2357

not the new technology itself, but rather using old models and old modes of thought in dealing with situations arising from new technology.

Privacy may mean different things to different people. Three different aspects of privacy are:

- **seclusion:** the desire to be left alone
- **property:** the desire to be paid for one's data
- **autonomy:** the ability to act freely

For individuals whose primary concern is seclusion, protection of their property will not suffice. For example, an individual who cares about seclusion will consider the receipt of any e-mail spam as a critical privacy violation, while one who cares about property may be satisfied to receive payment or discounts in exchange for receiving spam. Autonomy is an issue if people find their behavior is constrained by their concerns that their behavior is being tracked. A general definition that can capture most aspects of privacy is "the ability to control the dissemination and use of one's personal information."

We investigate the question of whether there are tradeoffs between privacy and other attributes, and if such tradeoffs are inherent. For example, there may be tradeoffs between security and privacy, as is frequently mentioned in the United States since the terrorist events of September 11, 2001. In order to protect national security, the argument goes, it is necessary to routinely perform authentication and identification of individuals, and to monitor their whereabouts and behavior, despite the privacy violation this creates. There are also potential tradeoffs between privacy and usability (as introducing privacy features may make systems more difficult to use), privacy and marketability (as customers may not be willing to pay extra for privacy-protecting solutions, and businesses may not be willing to give up collecting personal information they deem valuable), and even between different notions of privacy such as property vs. autonomy.

In order to hope to achieve privacy of data, several kinds of protection are needed. It is necessary to protect stored data, data in transit, and to have some control over the release of data. Protection of stored data and data in transit are well-studied and somewhat well-solved problems in the areas of computer security and network security, and are usually considered outside the scope of "privacy." Solutions usually involve the use of encryption and authentication to ensure that data is only sent to authorized parties and is only readable by those it is sent to. In contrast, most current privacy-oriented work, such as P3P and related tools [1], assumes data in transit and storage will be protected, and focuses on helping users to state their willingness to release data based on how that data will be used. However, there have been a number of cases where personal information was leaked in violation of a stated privacy policy due to a failure of computer security. In this sense, privacy requires security, and security violations can lead to privacy violations.

Current privacy-oriented solutions such as P3P deal primarily with the case of interaction between the *stakeholder*, whose personal data is involved, and an enterprise such as a business whose Web site is being visited. We call such data "transaction data." A second class of data is "authored data," which is created by the stakeholder(s). In this case, property is usually the relevant privacy issue. Digital rights management, which is geared toward guaranteeing the stakeholders' payments, aims to provide solutions. Both with transaction data and authored data, a common theme in many protection approaches is to label data in some way with the identities of the stakeholders and a description of the policies regarding use of the information.

A newer—and more difficult to control—class of data is "sensor data," or data that is collected by some kind of sensor. Some examples include:

- video surveillance cameras: the use of video surveillance cameras has become more and more pervasive, and such cameras have become smaller and more easily hidden or overlooked [3]. The privacy impact of surveillance cameras can be even greater if used in conjunction with face-recognition technology.
- various data mining applications related to national security or marketing, in which data is collected from diverse sources and combined in such a way as to reveal information about individuals' preferences, habits, or activities.
- desktop or keystroke monitoring software: such software is often used for intrusion or misbehavior detection in the workplace. Often workers may not be aware that they are being monitored in this way.
- GPS transmitters: for example, on taxicabs in order to help dispatchers provide better service. Unless protected properly by encryption, information from such transmitters can also be used, for example, for outside observers to determine the destinations of particular passengers.
- Radio-frequency identification (RFID) tags: it seems likely that in the near future, most products manufactured will contain an inexpensive RFID tag that broadcasts a unique 96-bit serial number when queried. Given that such products might include the clothing we wear and the money we carry and exchange, queries to RFID tags could potentially be used to track individuals' locations and interactions.
- wireless PDA's and other devices that broadcast recognizable identification information: for example, such services might allow a user to receive information from local businesses as she walks down the street. There is a clear tradeoff here between a user being open to services from entities they have no prior trust relationship with, and the potential for privacy invasions. Current implementations and standards tend to favor service provision over privacy.
- iris scans: it may be possible to do iris scans of individuals at reasonably large distances without their cooperation or knowledge. Such systems could be used, for example, to determine whether an individual has previously entered a building, even without necessarily identifying the individual.

Sensor data presents a real and growing privacy threat. In sensor data, the identities of the stakeholders are not necessarily clear at time of creation, nor are the identities of the data collectors or even the existence of the sensors necessarily known to the stakeholders. Hence, any privacy approach that labels data with the stakeholders at the time of creation in order to constrain later behavior appropriately cannot work. As the above examples make clear, this class of data is growing rapidly.

The privacy impact of huge amounts of sensor data could be tremendous, as sensor data often crosses the boundary between the "real world" and "cyberspace". We note that when sufficiently aggregated (particularly across multiple entities), transaction data can have many of the same properties as sensor data. It is for this reason that most privacy policies focus on when and how data will be shared with other parties. Relatedly, this is why people objected to systems such as DoubleClick's, that would track Web-site browsing history, even if only the IP address, rather than an individual's name, was associated with the transaction at the time of collection [2].

In many cases, product decisions by large companies or public organizations become de facto policy decisions as their products are widely adopted. Often such decisions are made without conscious thought or public discussion about the privacy impacts. This is particularly true in the United States where there is not a lot of relevant legislation regarding what is legal from a privacy perspective. As an example of this, consider the difference between the magnetic stripe cards used to pay for the Metro in Washington, DC and those in New York City. In the former case, card usage data is (reportedly) not stored on a per card basis, while in the latter, it is. These decisions were most likely made not on the basis of privacy, but rather because at the time of implementation in Washington, data storage and processing techniques were less advanced, while at the time of implementation in New York City, it was clear that data storage was possible and the processing might yield information that would help the transit system to run more efficiently. We note that the lack of privacy in New York City has been used with good outcome (for example, corroborating alibis of innocent suspects in criminal cases), as well as having the potential for bad uses. The point we wish to emphasize is not that one technology is obviously better than the other, but that important privacy-relevant decisions were made without public debate and awareness.

The main tradeoff for privacy is advancing technology which makes it possible to store and process large amounts of data. Even if it were possible to stop such technological advances, most people would not advocate this approach in order to protect privacy. Similarly, there are instances where health or security concerns can override privacy concerns. For example, if an unconscious person is admitted to a hospital emergency room, it is generally more important to allow the medical personnel access to the person's medical history than to maintain privacy of that information at all costs. Similarly, if there is a specific immediate terrorist threat, security concerns can temporarily override privacy concerns. Still, it is important not to blindly give up privacy in the name of security or other goals. For example, the privacy-invasive question "who is this person?" is not the same as the security-relevant question "is this a dangerous person?" Answering the former question instead of the latter both unnecessarily violates privacy and may miss some security-relevant threats, particularly in the case of first offenses. Regarding the tradeoff between privacy and usability, a good approach to trying to achieve both is a layered one involving reasonable defaults, easy and extensive customizations, and possibly visualization tools to help the user understand privacy-relevant attributes.

In many cases, the tradeoffs are to cost or power rather than an inherent conflict with privacy. That is, the apparent tradeoff between security and privacy may really be two tradeoffs: one between security and money, and the other between privacy and money. Similarly, the tradeoff between privacy and usability may in fact be a conflict between the power of technology providers or governments to provide and support solutions that do not provide both privacy and usability, and users willingness or need to use them.

In summary, while there is sometimes an inherent tradeoff between privacy and other attributes, it is important to realize that often it is possible to achieve other goals in conjunction with privacy.

References

1. L. F. Cranor and R. Wenning, "Why P3P is a Good Privacy Tool for Businesses and Consumers", *Gigalaw.com*, `http://www.gigalaw.com/articles/2002/cranor-2002-04.html`, 2002.
2. P. Jacobus, " "Cookies" targeted as Congress, advocates address Net privacy", *CNET News.com*, February 11, 2000.
3. A. Reeder, "To See and Be Seen", *The Atlantic Monthly*, Vol. 282, No. 1, p. 62, July 1998.