Beyond Secrecy: New Privacy Protection Strategies for the World Wide Web *

Daniel J. Weitzner

Decentralized Information Group
Principal Research Scientist
MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar Street, Cambridge, MA 02139
djweitzner@csail.mit.edu

In 1967, Alan Westin [1] set in motion the foundations of what most Western democracies now think of as privacy when he published his book, Privacy and Freedom. He defined privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." His careful collection of sociological, legal, and historical perspectives on privacy came at a time when people worried that human dignity would erode or that governments would tend toward tyranny, becoming tempted to misuse their newfound power over private data. Computer scientists shared these concerns. Following Westin's emphasis on privacy as confidentiality, much of the security and privacy research over the last four decades has concentrated on developing more and more robust access control and confidentiality mechanisms.

Today, despite the fact that technical innovation in cryptography and network security has enabled all manner of confidentiality control over the exposure of identity in information systems, the vast majority of Internet user remain deeply worried about their privacy rights and correctly believe that they are far more exposed today than they might have been a generation earlier. Have we just failed to deploy the proper security technology to protect privacy, are our laws inadequate to meet present day privacy threats, or is have business practices and social conventions simply rendered privacy dead? While there is some truth to each possibility, the central failure to achieve robust privacy in the information age can be traced to an a long-standing mis-identification of privacy with confidentiality and access control.

Privacy protection in an era in which information flows more freely than ever will require increased emphasis on laws that govern how we can use personal data, not just who can collect it or how long they can store it. Much of our current privacy views are based on controlling access to information. We believed that

* An earlier version of this talk appears in Weitzner, Daniel J., "Beyond Secrecy: New Privacy Protection Strategies for Open Information Spaces," IEEE Internet Computing , vol.11, no.5, pp.96-95, Sept.

if we could keep information about ourselves secret, prevent governments from accessing emails, and so on, then we would have privacy. In reality, privacy has always been about more than just confidentiality, and looking beyond secrecy as the sine qua non of privacy is especially important. New privacy laws should emphasize usage restrictions to guard against unfair discrimination based on personal information, even if it's publicly available. For instance, a prospective employer might be able to find a video of a job applicant entering an AIDS clinic or a mosque. Although the individual might have already made such facts public, new privacy protections would preclude the employer from making a hiring decision based on that information and attach real penalties for such abuses.

If we can no longer reliably achieve privacy policy goals by merely limiting access to information at one point on the Web, then what systems designs will support compliance with policy rules? Exercising control at one point in a large information space ignores the very real possibility that the same data is either available or inferable from somewhere else. Thus, we have to engineer Policy Aware systems based on design principles suitably robust for Web-scale information environments. Here we can learn from the design principles that enabled the Internet and the Web to function in a globally-coordinated fashion without having to rely on a single point of control. Colleagues at MIT, RPI, Yale and elsewhere are investigating designs for information systems that can track how organizations use personal information to encourage rules compliance and enable what we call information accountability, which pinpoints use that deviates from established rules [2, 3]. We should put computing power in the service of greater compliance with privacy rules, rather than simply allowing ever more powerful systems to be agents of intrusion.

Accountable systems must assist users in seeking answers to questions such as: Is this piece of data allowed to be used for a given purpose? Is a string of inferences permissible for use in a given context, depending on the provenance of the data and the applicable rules. Information accountability will emerge from the development of three basic capabilities: policy-aware audit logging, a policy language framework, and accountability reasoning tools. A policy-aware transaction log will initially resemble traditional network and database transaction logs, but also include data provenance, annotations about how the information was used, and what rules are known to be associated with that information. Cryptographic techniques will play an important role in Policy Aware systems, but unlike the current reliance of privacy designs today, cryptography will be more for the purpose of creating immutable audit logs and providing verifiable data provenance information, than for confidentiality or access control.

Access control and security techniques will remain vital to privacy protection – access control is important for protecting sensitive information and, above all, preserving anonymity. My colleague from UC Berkeley, Deirdre Mulligan, recounts a situation on the Berkeley campus in which a computer vision experiment on campus captured images of a group of female Iranian students engaged in a protest against Iranian human rights violations. Although they were free

from harm on the campus, the fact that the researchers recorded the images and made them publicly available on the project's Web site put the students' family members, many of whom were still in Iran, at grave risk. The department took down the images as soon as they realized the danger, but harm could have easily occurred already.

Clearly, the ability to remain anonymous, or at least unnoticed and unrecorded, can be vital to protect individuals against repressive governments. Although US law doesn't recognize a blanket right of anonymity, it does protect this right in specific contexts, especially where it safeguards political participation and freedom of association. Even though no general protection exists for anonymous speech, we have a right keep private our role in the electoral process. Courts will protect the right of anonymous affiliation with political groups, such as the NAACP, against government intrusion. Finally, of course, we don't want our financial records or sensitive health information spilled all over the Web.

Nevertheless, in many cases the data that can do us harm is out there for one reason or another. With usage restrictions established in law and supported by technology, people can be assured that even though their lives are that much more transparent, powerful institutions must still respect boundaries that exist to preserve our individual dignity and assure a healthy civil society.

References

1. Westin, A.: Privacy and Freedom. The Bodley Head (1967)
2. Weitzner, Abelson, Berners-Lee, Feigenbaum, Hendler, Sussman: Information Accountability. Technical Report MIT-CSAIL-TR-2007, MIT (2007)
3. Hanson, Kagal, Sussman, Berners-Lee, Weitzner: Data-Purpose Algebra: Modeling Data Usage Policies. In: IEEE Workshop on Policy for Distributed Systems and Networks (POLICY). (2007)