

Privacy and Security

Privacy Research Directions

What must we learn in order to support privacy requirements as technology advances?

NOT SINCE THE early 1970s, when computing pioneer Willis Ware chaired the committee that produced the initial Fair Information Practice Principles,¹⁰ has privacy been so much in the U.S. public eye. Edward Snowden's revelations, as well as a growing awareness that merely living our lives seems to generate an expanding "digital exhaust," have triggered many workshops and meetings.^{1,5,11,12} An alphabet soup of advisory groups—PRG,^a PCLOB,^b PCAST^c—have produced privacy-related reports.^{2,6,7-9} The wheels are turning at NITRD^d to produce a national strategy for privacy research, perhaps paralleling the federal strategy for cybersecurity research and development.³ I have participated in a number of these and have developed my own view of privacy and privacy research. My U.S. perspective may differ from those from different backgrounds; privacy views vary with culture.

Some characterize privacy in terms of harms: to have suffered a loss of privacy that is actionable, there must be some way to characterize the harm



an individual suffers as a result of the privacy breach. This practical view motivates many privacy concerns: the data revealed may cause the loss of a benefit or service. However, this view runs into trouble where the damage seems primarily psychological—I really do not want my neighbor to know if I have unconventional sexual practices or have had cosmetic surgery, and I may suffer psychologically or emotion-

ally from such revelations, but it may be difficult to characterize the loss in a way that can be compensated. Further, it may be difficult to know that a harm has occurred—I may be deprived of an opportunity to be employed by the disclosure of private information through a breach I am unaware of.

Domain-specific privacy definitions and rules may be needed. I once preferred the uniform structure of privacy

- a President's Review Group on Intelligence and Communications Technologies: <http://1.usa.gov/1dY0nmm>
- b Privacy and Civil Liberties Oversight Board: <http://bit.ly/1NSRto6>
- c President's Council of Advisors on Science and Technology: <http://1.usa.gov/1NSRxnJ>
- d National Information Technology Research and Development program office: <http://1.usa.gov/1QB4W97>

Privacy Observations

- ▶ “Privacy” has many definitions. Perhaps most stringent is privacy as the right to control fully the flow of one’s personal data. But only those willing to forego most modern communication, transportation, and payment systems could today approach such an objective.
- ▶ Even people who say “I have nothing to hide” still value privacy. What they usually mean is “I have nothing to hide from law enforcement or others with legal authority to examine my records.” They do not mean, “I have nothing to hide from a parent, sibling, child, neighbor, blogger, or journalist.”
- ▶ The 40-year-old Fair Information Practice Principles remain sensible, but the average citizen has felt their effect mostly in paperwork generated by the notice and consent provisions.
- ▶ The schemes used to inform people about privacy and gain their consent are ineffective. The increasing ubiquity of automated sensing makes them more so. At one recent workshop, no panelist or audience member was willing to defend current “notice and consent” mechanisms.
- ▶ Government and private surveillance mechanisms generate substantial quantities of data without engaging the subjects under surveillance. Consent or even notice seems infeasible.
- ▶ Mass surveillance pits expected gains in national security for society as a whole against privacy of individuals, an unequal comparison. Mass surveillance affects society. Endorsing auditing and oversight mechanisms to enforce “good behavior,” from video monitoring to catch fraud in stores to Inspectors General for Federal agencies, suggests individuals—and by extension, society—behave differently when knowingly watched.
- ▶ Anonymization of data can help preserve privacy, but it can also limit the benefits gained from analyzing a dataset, while remaining vulnerable to serious attempts at re-identification.

regulations of the European Union’s Data Protection Directive to the U.S. patchwork of laws and regulations that separately cover health records, educational records, legal proceedings, business transactions, and so on. Now I am less sure. One of the definitions of privacy that continues to seem useful to me is that privacy is preserved when information remains within a specified context—financial information stays with my financial advisor or broker, religious information with my religious counselor, health information with my medical practice, educational information with my school, and so on.⁴ Perhaps it is better to continue to use policies that take these contexts and the semantics of the information into account and to strive for, but not insist on, unification. In this case, it is also necessary to specify when information can be allowed to move between normally isolated contexts, for example to deal with an emergency.

A legal regime in which data “belongs” to an owner who then has complete dominion over it is too simple to accommodate future needs both for preservation of useful notions of

privacy and for productive use of data. The fact that a private party engages in a transaction with a business or public service, be it a grocery purchase, telephone call, email message, or database query, should not necessarily entitle the business or service to unlimited use or publication of the data from that transaction. In fact, there are already many cases where there are competing interests in particular data and the custodian of data does not exercise complete control over it. In the future, it may make sense for data custodians to be bound by “responsible use” policies that depend on how the data was collected, the domain of data collected, and other factors. Constraints need to accompany data in a form that enables the recipient to continue to enforce them easily.

To gain the benefits of having large datasets to analyze (perhaps most apparent in healthcare, but in many other domains as well), anonymization and differential privacy will be of some, but limited, use. It will be essential for the subjects whose data is collected to trust that the custodians of their data will handle it properly, in

accordance with whatever responsible use policies are established for it. Otherwise the subjects may withhold their data and the societal benefits will be lost. Because humans and systems are fallible, there will undoubtedly be some instances where sensitive data is lost or mishandled. To maintain public trust in the face of such incidents it will be important to assure data subjects that the custodians can be held to account: mechanisms must be provided that enable injured parties to detect misuse and obtain redress.

Potential Areas of Research

With those thoughts in mind, I offer some potential privacy-related research areas, in no particular order, and with some overlap.

Effective privacy definitions, and in particular, domain-specific definitions of privacy. HIPAA (internationally considered a strongly protective model) and other regulations already provide what might be considered domain-specific rules for data privacy in healthcare. These have received some research attention. Other domains, including law enforcement, finance, and intelligence, might benefit from efforts to characterize the data involved, how it should be handled within the domain of use, and under what conditions and controls it might be allowed to flow to different domains.

Effects of surveillance on social groups of different sizes. For example, has the chilling effect of surveillance on free expression been studied systematically? I am not a social scientist, so I may be unaware of research in this area, but research results, if they exist, must be aired, and if they do not exist, they deserve study.

Development of better languages for specifying privacy policies. Not a new area of research, perhaps, but effective solutions do not seem to be available. Languages are needed that enable specification of policies that can be enforced algorithmically and also that can be understood by the public.

Accountability mechanisms for privacy violations. Both detection of a privacy violation and the ability to trace the violation back to responsible individuals are important. Detecting violations will of course imply there is an expressed policy that is to be enforced. Tracing violations has implications for

authentication and auditing. Some financial systems incorporate accountability mechanisms for purposes of fraud deterrence and detection, and some health records systems incorporate mechanisms to account for privacy violations, but these mechanisms will need to find homes in a much broader range of systems with various privacy policy enforcement needs. The ability to support provenance for both data and programs at scale is needed.

Techniques and mechanisms for tracking information flow. To me, the fundamental nature of a privacy violation is an improper information flow. Privacy policy needs to distinguish proper and improper flows and to enable authorization of exceptions. Capabilities for tracking the flow of information within programs have matured substantially in the past decade. Those capabilities need to be extended to systems of programs.


Techniques for binding policies to data and enabling distributed enforcement. If the data itself can carry the policy to be enforced along with it, each domain in which it appears can apply appropriate enforcement. One might imagine the data also collecting an audit trail as it moves from place to place. Cryptographic approaches may help.

Techniques for identifying and quantifying benefits of large-scale data analysis and costs of privacy harms. It is a tall order to model in advance the benefit one may gain from analyzing some large dataset, since one does not know what might be learned. On the other hand, the analysis is usually undertaken with some objective in mind, and it might be possible to quantify what is to be gained if that objective is realized. Similarly, some resources need to be devoted to anticipating privacy harms and what damages may occur if large datasets are abused. These kinds of trade-offs must be understood as well as possible at the time people are deciding whether or not to initiate new projects if there is to be any rigorous risk/benefit analysis in this sphere.

Conclusion

Privacy may be difficult to define and culturally dependent, but it nevertheless seems to be universally valued. Future computing systems must in-

Privacy may be difficult to define and culturally dependent, but it nevertheless seems to be universally valued.

corporate mechanisms for preserving whatever privacy policies people and societies decide to embrace, and research is needed to identify those mechanisms and how they may best be applied. 

References

1. National Academy of Science Raymond and Beverly Sackler U.S.-U.K. Scientific Forum on Cybersecurity, Dec. 8-9, 2014, Washington, D.C.
2. Networking and Information Technology Research and Development (NITRD) Program. Report on Privacy Research Within NITRD. (Apr. 2014); <http://1.usa.gov/1IUfakz>
3. Networking and Information Technology Research and Development (NITRD) Program. Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program. (Dec. 2011); <http://1.usa.gov/1NgEUFN>
4. Nissenbaum, H. Privacy as contextual integrity. *Washington Law Review* 79 (2004), 119-158.
5. 2015 NSF Secure and Trustworthy Cyberspace PI meeting (Jan. 5-7, 2015), Washington, D.C.; <http://bit.ly/10XJxVV>
6. President's Review Group on Communications and Intelligence Technologies. *Liberty and Security in a Changing World*. (Dec. 12, 2013); <http://1.usa.gov/1cBct0k>
7. President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective*. (May 2014); <http://1.usa.gov/1rTipM2>
8. Privacy and Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of the Foreign Intelligence Surveillance Act*, July 2, 2014; <http://bit.ly/1FJat9g>
9. Privacy and Civil Liberties Oversight Board, *Report on the Telephone Records Program Conducted under Section 215 of the USA PATRIOT Act and on the Operations of the Foreign Intelligence Surveillance Court*. (Jan. 23, 2014); <http://bit.ly/1SRiPke>
10. *Records, Computers, and the Rights of Citizens: Report of the Secretary's Committee on Automated Personal Data Systems* (July 1973). Dept. of Health, Education and Welfare, DHEW(OS), 73-94; <http://1.usa.gov/1RIZLom>
11. Second Annual CATO Surveillance Conference (Oct. 21, 2015), Washington, D.C.; <http://bit.ly/1MEsY05>
12. U.S. Privacy and Civil Liberties Oversight Board (PCLOB) Workshop "Defining Privacy", Nov. 12, 2014, Washington, D.C.; <http://bit.ly/1ReOmgT>

Carl Landwehr (carl.landwehr@gmail.com) is Lead Research Scientist the Cyber Security Policy and Research Institute (CSPRI) at George Washington University in Washington, D.C., and Visiting McDevitt Professor of Computer Science at LeMoyne College in Syracuse, NY.

Copyright held by author.

Calendar of Events

February 14-17

TEI '16: 10th International Conference on Tangible, Embedded, and Embodied Interaction, Eindhoven, the Netherlands, Sponsored: ACM/SIG, Contact: Saskia Bakker, Email: s.bakker@tue.nl

February 21-23

FPGA'16: The 2016 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, Monterey, CA, Sponsored: ACM/SIG, Contact: Deming Chen, Email: dchen@illinois.edu

February 22-25

WSDM 2016: 9th ACM International Conference on Web Search and Data Mining, San Francisco, CA, Sponsored: ACM/SIG, Contact: Paul N. Bennett, Email: paul.n.bennett@microsoft.com

February 27-March 2

CSCW '16: Computer Supported Cooperative Work and Social Computing, San Francisco, CA, Sponsored: ACM/SIG, Contact: Meredith Ringel Morris, Email: merrie@microsoft.com

March

March 1

I3D '16: Symposium on Interactive 3D Graphics and Games, Sponsored: ACM/SIG, Contact: Chris Wyman, Email: chris.wyman@ieee.org

March 2-5

SIGCSE '16: The 47th ACM Technical Symposium on Computing Science Education Memphis, TN, Sponsored: ACM/SIG, Contact: Jodi L. Tims, Email: jltims@bw.edu

March 7-10

IUI'16: 21st International Conference on Intelligent User Interfaces, Sonoma, CA, Sponsored: ACM/SIG, Contact: John O'Donovan, Email: jodmail@gmail.com