Educating Engineers

Teaching Privacy in a World of Open Doors

Susan Landau

Two decades ago, the number of people thinking hard about electronic privacy roughly mirrored the attendance of the annual Computers, Freedom, and Privacy Conference—about 250 people. That sounds hard to believe now. Companies have Chief Privacy Officers, law schools have information privacy courses, and Internet companies have privacy counsels. The International Association of Privacy Professionals, which certifies privacy professionals and privacy managers, was organized in 2000; it now has 14,000 members. Its members work within organizations to assure "data protection, information auditing, information security, legal compliance and/or risk management" (www.privacyassociation.org/about_iapp/mission_and_background).

But privacy isn't only about compliance, it's also about creating systems that collect information while designing them to ensure they protect privacy. To put this another way, privacy requires insight not only from lawyers and auditors but also from the social scientists, computer scientists, and engineers who design and build the systems that touch and consume people's data.

How are the universities doing in teaching this endeavor? Not so well. Privacy is inherently multidisciplinary, involving computer science, law, policy, social science, humanities, and design. It involves thinking about how people think and make choices and how technologies work. It's about technology—cryptography and controls—and values.

Such a mix is quite challenging for academic institutions. A handful of institutions have privacy courses in their computer science curriculum. One, Carnegie Mellon University, has a new master's program in privacy engineering.[1] But for engineers encountering privacy issues in the context of building a system, learning about privacy is largely ad hoc—if it occurs at all. Considering how critical privacy has become in a world of online social networks, ubiquitous communication devices, and the daily threats to privacy, we should be doing much better. This column is about teaching privacy. The subject is rich, and multiple approaches exist. The one I take here is based on my and several colleagues' experience.

Although I focus here on teaching privacy to computer scientists, I want to first mention the law-school approach, which is, of course, lawyerly. The topics in a typical law school course on information privacy include the development of privacy within the law; privacy law in commercial practice, health information, and communications; privacy and data protection, including the international aspects of this; and regulatory frameworks for privacy. In rare cases, mostly those in which the faculty member does cyberlaw research, the course might cover technological protections for privacy.

Undergraduate and graduate computer science courses in privacy have different audiences and different goals from law school ones; they also differ from each other. An undergraduate course should

1540-7993/14/$31.00 © 2014 IEEE

present myriad privacy approaches, whereas a graduate course might well focus on current technological research.

Teaching Undergraduates

Undergraduate education is about teaching students to think. This is particularly challenging when the topic is privacy, about which everyone has an opinion and few agree. The bigger challenge, however, in teaching privacy is the breadth of the field, which draws on many fields, including philosophy, history, law, psychology, anthropology, and technology.

An undergraduate privacy course could focus solely on technical aspects (cryptography, anonymity techniques, security, and so on). However, teaching privacy to undergraduates affords an opportunity to demonstrate how social choices inform technical decisions. An undergraduate course focused narrowly on technology misses educating computer science students about law, policy, design, and values. This is something we should be doing more of in a world where computers affect almost every aspect of human endeavors. For a computer science professor, a privacy course's breadth of topics presents a serious challenge. Another large challenge is that privacy approaches are culturally dependent; some discussion of this (for example, the distinctions between US and European privacy approaches) is appropriate.

Pull quote for departments: Approximately 20-25 words. Pull quote for departments: Approximately 20-25 words. Pull quote for departments: Approximately 20-25 words.

A privacy course should aim to present sufficient information and context such that students can argue intelligently about privacy. A second aim should be to include diverse viewpoints: teach the course so that it brings in students from other disciplines, including the social sciences, humanities, and arts. This presents yet more challenges, for although privacy is a social construct, in computer science, it's also a technical one.

Consider the confidentiality assurances of cryptography or perfect forward secrecy, the anonymity provided by Tor (www.torproject.org), or the privacy challenges of single sign-on. These topics can easily overwhelm nonmajors. So, tradeoffs exist. An undergraduate privacy course that has noncomputer-science students should present public-key cryptography, although probably not at the level of detail that explains the minutiae of potential attacks (and which exponents to avoid in an RSA implementation). A course that includes nontechnical students gains a richness of discussion that makes such a tradeoff well worth it. Fortunately, there are sources of technical material presented at levels accessible to students who aren't scientists or engineers. (These include "A Visual History of Cryptography and Encryption,"[2] chapter 2 of *Privacy on the Line: The Politics of Wiretapping and Encryption*,[3] chapter 3 of *Surveillance or Security? The Risks Posed by New Wiretapping Technologies*,[4] and the initial material in chapters 3, 4, and 5 of *Security Engineering: A Guide to Building Dependable Distributed Systems*.[5])

Although I'm generally reluctant to be prescriptive in course content, I believe that a privacy course must begin by discussing, Why privacy? By that, I mean a real discussion of such issues as, Why should I care about privacy if I have nothing to hide?[6] Wouldn't we all be better off if all information was always recorded and visible? Lecture courses aren't amenable to group discussion, but there are various well-known ways to break out of such restrictions. My preferred way to engender this initial discussion is to show *The Lives of Others*, a superb movie about Stasi surveillance in East Germany and the ensuing corruption of society. This movie opens students' eyes in a way that few other approaches can.

Because privacy draws on such varied sources, I develop a vocabulary and context early on. Useful sources include Samuel Warren and Louis Brandeis's classic 1890 *Harvard Law Review* article on privacy,[7] the *Stanford Encyclopedia of Philosophy* definition of privacy,[8] Daniel Solove's privacy taxonomy,[9] and Helen Nissenbaum's contextual-privacy paper.[10] Others take different approaches to the course; see, for example, "Privacy Technologies: An Annotated Syllabus."[11]

Legal issues create a framework from which all else follows. The Fair Information Practice Principles (http://epic.org/privacy/consumer/code_fair_info.html) are dated in some ways, but they're nonetheless the backbone for US and European privacy regulation. Studying these principles and their genesis provides a crucial understanding of where we are in privacy and where we might need to go.

Where you're teaching will matter because the legal foundation for privacy will differ. A US-based course, for example, would want to examine the basis of privacy in the Bill of Rights (the First, Third, Fourth, and Fifth Amendments) and subsequent legal interpretations. European courses might well want to focus on the European Declaration of Human Rights, the Organization for Economic Development and Cooperation's Privacy Principles (http://oecdprivacy.

org), and so on. Discussing both perspectives in a single course provides contrast—an excellent paper here is "Privacy on the Books and on the Ground"[12]—and informs computer science students about the differing standards to which they might be building systems.

With the legal framework in place, you can proceed in various directions. A privacy course should cover both technologies and threats; valid arguments exist for covering these in either order. On the technical side, cryptography and anonymization tools are important. But that's just scratching the surface; there's a wealth of material, including work in differential privacy,[13] privacy design in identity management,[14] and decentralized architectures. The course should also cover technical material on failure, ranging from failure such as de-anonymization of anonymized sets[15] to more social failures—for example, the failures of privacy notices and the complexity of privacy decision making in online social networks. Choices must be made on the basis of the course's focus and the depth to which the course can cover technical material (for example, the detail of differential privacy is too difficult for an undergraduate course for nonscientists). Is it a technical course with some social-science content or a social-science course that uses the technical material to provide appropriate background?

The course should also cover current privacy threats. Here, the best sources are often journalistic; separating wheat from chaff is paramount (indeed, one purpose of an undergraduate privacy course should be to enable students to critically read such stories). The *Wall Street Journal* has had an excellent series of articles on online privacy.[16–18] Other important sources come from the economics of information privacy; I recommend "Privacy and Rationality in Individual Decision Making"[19] and papers from the annual Workshop on the Economics of Information Security (WEIS).

Finally a number of "special" topics are appropriate for a privacy course. This includes the use of closed-circuit TV[20] and wiretapping for investigations,[3] the use of genomic data (anonomymized or otherwise), the privacy risks raised by the Internet of Things, and the right to read anonymously.[21] Such topics can be directly part of the course or can provide an excellent source for student papers and projects.

Pull quote for departments: Approximately 20-25 words. Pull quote for departments: Approximately 20-25 words. Pull quote for departments: Approximately 20-25 words.

The goal should be to teach the undergrad computer science student—or the literature or sociology major—to ask probing questions when using a service or new technology or helping to create one. What information is being collected? Is the information being shared with others? Does the user have control over release to third parties? What's the consequence of not supplying the requested information?

Students should learn to question the role of technological design decisions. Which ones create privacy problems, and which ones address them? Which solutions are usable?[22] You can teach this in many ways, including student projects. For example, students could investigate what users think they're doing when they specify privacy settings in social networks[23] or examine the usability of privacy protection tools such as Tor. Students should develop a nuanced appreciation of privacy (for example, the differences between it and anonymity). They should also broaden their view of privacy to include awareness of it in different cultures and how it changes over time.

Such a broad course is challenging to students and perhaps even more challenging to the professor teaching it. "A Critical Review of 10 Years of Privacy Technology"[24] and *Engaging Privacy and Information Technology in a Digital Age*[25] might prove useful.

Colleges and universities are, at least in theory, amenable to interdisciplinary work. Examples of such courses include Paul Ohm's Technology of Privacy course at the University of Colorado Law School and an upper-level undergraduate course Jim Waldo taught for a number of years at Harvard. Teaching the broad course I just described can be made easier by inviting guest speakers from other disciplines (including law, policy, information science, and privacy researchers from other institutions) or coteaching the course with a colleague from another discipline. The latter, complicated in some ways, can be quite enriching and might even lead to research collaborations. But the best part of teaching a course so wide in scope is the education you give and get in the field.

Teaching Graduate Students

Starting with focus, a graduate privacy course is a different kettle of fish than an undergraduate one. Although an information science program might contain a privacy course—in which case it could be a first- or second-year graduate course—a graduate privacy course in a computer science department typically is taught by a privacy researcher (this is not necessarily the case for an undergraduate course). So, the graduate course tends to emphasize technical results in privacy and not privacy's social aspects—that is, law, regulation, economics, psychology (including human–computer interaction [HCI]), and so on. The material that forms a third of the undergraduate course becomes the main body of the graduate one (covered, of course, in greater depth).

That said, non-computer scientists might be taking the course. Unlike an undergraduate course in which the students all cover the same material, in a graduate course, different students might take alternative paths emphasizing different material. One benefit is that law, policy, or social-science students attending a technologically oriented graduate course can present privacy-related material from the domains in which they're expert.

With the technical aspects in mind, topics should include cryptography and its failures in deployment, data collection and tracking (including using metadata and third-party data collection to develop user profiles), anonymization tools (pseudonymity, Tor, *k*-anonymity, and differential privacy), and attacks on and failures of those tools. Researchers should consider how people make privacy choices[19] and the economic tradeoffs arising from "free" services and targeted advertising.[26] The WEIS papers and the research presented at the Symposium on Usable Privacy and Security are useful. Because this type of research might involve people, the course should spend some time, even if brief, covering appropriate experimental design and the requirements of human-subject research.

Studying one system in depth is useful, whether it's the privacy controls of smart phones and what data is "shared" from them, the design choices in developing a federated identity management system, or the complexities of developing a do-not-track system. When students do so, the choices—simplicity of notice versus full explication, pseudonymous sharing of identities, and the differing interests of the players involved—become clearer. Students develop a better sense of the tradeoffs involved in designing for privacy.

An undergraduate course aims to teach the richness of privacy; a graduate course should do that and foster an appreciation of the differences between privacy and security. Many techniques, including cryptography and HCI design of access controls, are the same for both privacy and security but have different purposes. This privacy aspect is more akin to using a lock on a bathroom door than to having one on the front door. The bathroom lock signals "stay out," while also being good enough to keep out a prying five-year-old (but not a determined burglar). Developing students' sense of privacy needs and tradeoffs should be a graduate course's fundamental goal.

Privacy through the Curriculum

Privacy runs through all human endeavors: personal relationships; roles and relationships at work; and interactions with government, stores, friends on the soccer field, and acquaintances at church or the market. Privacy similarly runs through many aspects of the technology we develop. Yet when a computer science department offers a privacy course, it's an elective. That's probably the correct call for now, but it also provides an opportunity. Teaching privacy in a separate course affords the ability to spend time on techniques and technologies, social and policy aspects, and laws and regulations of privacy. However, we should be teaching students about privacy through all aspects of the technologies they'll develop. That's an argument for teaching privacy design in database, networking, security, and hardware courses.

For each of these technologies, discuss privacy and how to build privacy-protective designs. For databases, this could be work on searching within encrypted databases. For networks, it might be the privacy issues raised by sensor networks. For security, myriad issues exist, including the differences (and similarities) between a security solution and a privacy one. The first programming course should also cover privacy issues.

We should be showing students that system design should consider privacy, and there's no better way to do so than within the context of building systems. So, although privacy courses are important and have their place, they should complement the teaching of privacy across the curriculum. If we are to preserve any form of privacy in society, we must build an appreciation of it into computer science courses at multiple levels. ■

Acknowledgments

I thank Arvind Narayanan, Jim Waldo, and especially Seda Gurses; their suggestions improved this article.

References

1. L.F. Cranor and N. Sadeh, "A Shortage of Privacy Engineers," *IEEE Security & Privacy*, vol. 11, no. 2, 2013, pp. 77–79.
2. M. Felker, "A Visual History of Cryptography and Encryption," *Tom's IT Pro*, 2012; www.tomsitpro.com/articles/encryption-cryptography-ceasear_cipher-RSA,5-24-17.html.
3. W. Diffie and S. Landau, *Privacy on the Line: The Politics of Wiretapping and Encryption*, revised ed., MIT Press, 2007.
4. S. Landau, *Surveillance or Security? The Risks Posed by New Wiretapping*

Technologies, MIT Press, 2011.

5. R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*, John Wiley & Sons, 2008.
6. D. Solove, "'I've Got Nothing to Hide' and Other Misunderstandings of Privacy," *San Diego Law Rev.*, vol. 44, 2007, pp. 745–772.
7. S. Warren and L. Brandeis, "The Right to Privacy," *Harvard Law Rev.*, vol. 4, no. 5, 1890.
8. "Privacy," *Stanford Encyclopedia of Philosophy*, 2013; http://plato.stanford.edu/entries/privacy.
9. D. Solove, "A Taxonomy of Privacy," *Univ. of Pennsylvania Law Rev.*, vol. 154, no. 3, 2006, pp. 477–560.
10. H. Nissenbaum, "A Contextual Approach to Privacy Online," *Daedalus*, vol. 140, no. 4, 2011, pp. 32–488.
11. Narayanan, "Privacy Technologies: An Annotated Syllabus," *Proc. Privacy Enhancing Technologies Symp.*, 2013.
12. K.A. Bamberger and D.K. Mulligan, "Privacy on the Books and on the Ground," *Stanford Law Rev.*, vol. 63, 2010, pp. 247–316.
13. C. Dwork, "A Firm Foundation for Private Data Analysis," *Comm. ACM*, vol. 54, no. 1, 2011, pp. 86–95.
14. E. Birrell and F. Schneider, "Federated Identity-Management Systems: A Privacy-Based Characterization," *IEEE Security & Privacy*, vol. 11, no. 5, 2013, pp. 36–48.
15. A. Narayanan and V. Shmatikov, "Robust De-anonymization of Large Sparse Datasets," *Proc. 2008 IEEE Symp. Security and Privacy*, 2008, pp. 111–125.
16. J. Angwin, "The Web's New Gold Mine: Your Secrets," *Wall Street J.*, 30 July 2010.
17. E. Steel and J. Angwin, "On the Web's Cutting Edge, Anonymity in Name Only," *Wall Street J.*, 4 Aug. 2010.
18. J. Scheck, "Stalkers Exploit Cellphone GPS," *Wall Street J.*, 3 Aug. 2010.
19. A. Acquisti and J. Grossklags, "Privacy and Rationality in Individual Decision Making," *IEEE Security & Privacy*, vol. 3, no. 1, 2005, pp. 26–33.
20. A. Sasse, "Not Seeing the Crime for the Cameras," *Comm. ACM*, vol. 53, no. 2, 2010, pp. 22–25.
21. J. Cohen, "A Right to Read Anonymously: A Closer Look at 'Copyright Management' in Cyberspace," *Connecticut Law Rev.*, vol. 28, 1996, pp. 981–1039.
22. A. Whitten and J.D. Tygar, "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0," *Proc. 8th Usenix Security Symp.*, McGraw-Hill, 1999.
23. M. Madjski, M. Johnson, and S.M. Bellovin, "A Study of Privacy Settings Errors in an Online Social Network," *Proc. 4th IEEE Int'l Workshop Security and Social Networking*, 2012, pp. 340–345.
24. G. Danezis and S. Gurses, "A Critical Review of 10 Years of Privacy Technology," *Proc. 4th Bi-annual Surveillance Cultures Conf.*, 2010.
25. J. Waldo, H. Lin, and L. Miller, *Engaging Privacy and Information Technology in a Digital Age*, Nat'l Academies Press, 2007.
26. J.R. Mayer and J.C. Mitchell, "Third-Party Web Tracking: Policy and Technology," *Proc. 2013 IEEE Symp. Security and Privacy*, 2013, pp. 413–427.

Susan Landau is the author of *Surveillance or Security? The Risks Posed by Wiretapping Technologies* (MIT Press, 2011) and coauthor of *Privacy on the Line: The Politics of Wiretapping and Encryption* (MIT Press, revised ed., 2007). She taught about privacy at Wesleyan University in the 1980s and taught a freshman seminar on privacy at Harvard in 2012. Contact her at susan.landau@privacyink.org.