Privacy Engineering: Shaping an Emerging Field of Research and Practice

Seda Gürses | Princeton University
Jose M. del Alamo | Universidad Politécnica de Madrid

The emerging field of privacy engineering responds to the gap between research and practice, systematizing and evaluating approaches to capture and address privacy issues while engineering information systems.

Privacy engineering is an emerging research framework that focuses on designing, implementing, adapting, and evaluating theories, methods, techniques, and tools to systematically capture and address privacy issues in the development of sociotechnical systems.

We primarily situate the field in software engineering yet expect it to build on an intradisciplinary foundation, leveraging techniques and tools from various computer science subdisciplines, such as security engineering, human–computer interaction, and machine learning. Because law, societal norms, ethical conceptualizations, and technological advances inform privacy, the field is also inevitably interdisciplinary. Furthermore, developing a robust practice will benefit from knowledge of existing business practices as well as organizational studies, and psychology. Finally, we expect legislative, policy, and organizational schemes to play a role in incentivizing the development and adoption of privacy engineering in practice; these also require evaluation through an engineering-centric lens.

Attention to privacy engineering as a research topic increased dramatically after 2012 (see Figure 1). To facilitate the development of this emerging field, we organized the First International Workshop on Privacy Engineering (IWPE), co-located with 36th IEEE Symposium on Security and Privacy. IWPE provides a forum for those interested in tackling the gaps and challenges in privacy engineering. With its explicit focus on engineering techniques and its interdisciplinary program committee with members from computer science, law, policy, social sciences, humanities, and design, the workshop complements existing venues that focus mainly on presenting privacy solutions, like the Symposium on Usable Privacy and Security (https://cups.cs.cmu.edu/soups) or treat privacy as a subfield of security engineering, like the Privacy Enhancing Technologies Symposium (https://petsymposium.org).

The first iteration of the workshop attracted 47 delegates from academia, industry, government, and civil society. The presentations introduced different models and frameworks for understanding privacy; illustrated several methods, techniques, and tools; and provided case studies of privacy-engineering practice in enterprise systems. The programs and presentations can be found at the workshop website, http://ieee-security.org/TC/SPW2015/IWPE.

The Need for Privacy Engineering
Privacy research in computer science has produced a rich array of privacy solutions; however, the
integration of these into everyday engineering practice has been slow. In recent years, reports of privacy violations and technology companies’ failure to fulfill basic data protection requirements have become commonplace, suggesting that we’re far from applying privacy design know-how in practice. The consequences are most evident in the exorbitant number of data breaches: in the US alone, 4,700 breaches have been made public since 2005.2

But when it comes to privacy, a data breach is only one concern among many. Subtle engineering decisions that ignore users’ privacy needs might have far-reaching consequences. Recent highlights include Snapchat violating user expectations and privacy by not deleting users’ messages; Firefox extension NoScript’s defaults leading to deanonymization attacks on Tor users; and Facebook apps allowing the sharing of users’ friend networks with advertisers. Moreover, when these design decisions concern global infrastructures, such as cloud services, grids, and mobile networks, privacy protections applied at higher layers might be rendered moot. Past reports of Apple, Google, and Microsoft collecting location information gathered by their respective mobile devices from Wi-Fi hotspots—even when users turn off location tracking—and Snowden’s revelations about the US National Security Agency and the UK Government Communications Headquarters surveillance programs illustrate such domino effects.3

The different examples underscore that addressing privacy is relevant when engineering technical infrastructures, implementing organizational controls, and designing user experience (UX). They also imply that privacy solutions are potentially unknown to engineering teams, not practical to integrate into engineering activities, not of interest to the organizations, or nonexistent. Which of these cases hold and when? And what would it take to facilitate an engineering practice that addresses these issues? These remain open questions.

Researchers and practitioners who engage in the topic have made groundbreaking contributions to the field but have rarely attended to the development of a privacy-engineering practice. Their contributions include a wide array of technical solutions that help protect users’ privacy from diverse adversaries and in different social contexts.4 These solutions are informed by rigorous investigations into particular ways in which new technologies threaten privacy.5,6 They also illustrate the way experts successfully utilize or transform techniques from various computer science subfields, such as information security, software engineering, and social computing.7 However, few of these efforts are invested in systematizing or generalizing their approaches so other organizations and engineers can adopt and integrate them into their daily practices. In the research community, some have proposed monolithic privacy-engineering methods;8,9 however, these tend to assume a one-size-fits-all approach that disregards context, such as organization type and development practices, imminent privacy threats, and informational norms.

Motivating a New Field of Study
Privacy engineering addresses the lack of generalization in existing approaches; shortage in efforts to integrate different subdisciplines’ techniques and tools; the need to evaluate proposed approaches in different social, organizational, technical, and legal contexts; and concrete challenges emerging from the evolution of engineering practices, technical architectures, legal frameworks, and social expectations. Included in this research framework are projects that critically assess ways to respond to regulators’ and organizations’ increasing demands to implement and settle policy “through architecture, configuration, interfaces, and default settings,”10 also called privacy or data protection by design.11,12 The field intends to address these gaps by consolidating existing privacy research.

Over the past several decades, computer scientists have recognized the quest to build privacy-friendly systems as a research challenge. Most efforts have followed three prominent approaches. The first is what Sarah Spiekermann and Lorrie F. Cranor identified as privacy by architecture,9 which aims to minimize the collection or inference of sensitive information by unintended parties, typically service providers. Researchers develop technologies that enhance privacy by applying techniques that hard-code constraints on data collection and processing in systems and by ensuring that no entity can single-handedly undo these constraints. Privacy-enhancing technologies (PETs), such as Tor for anonymous communications or private information retrieval protocols for confidential search, are developed
using this approach. PETs can be used as stand-alone privacy technologies, like Tor, or function as the primitives in a privacy engineer’s toolbox, as in the case of zero-knowledge proofs and differential privacy.

Spiekermann and Cranor identified a second approach as privacy by policy.8 It aims at “protecting consumer data from accidental disclosure or misuses and facilitating informed choice options”—in other words, enforcing measures to ensure compliance with principles of data protection laws in information systems. Depending on jurisdiction, these requirements might include specifying and notifying users of the purpose of collection; limiting collection and use to this purpose; being transparent about additional recipients of the data; and providing users access to their data for verification, correction, and deletion. Proposed technologies include policy specification languages, policy negotiation and enforcement mechanisms, and design techniques to improve the readability of privacy policies.

A third approach, let’s call it privacy by interaction, focuses on sociotechnical designs that would improve users’ agency with respect to privacy in social settings. The approach captures privacy matters that arise, for example, between peers or in a workplace due to the introduction of information systems. These “lateral privacy” concerns are related to but often distinct from concerns regarding organizations collecting and processing data, as privacy-by-policy approaches address, and unintended inferences, as privacy by architecture tackles. The social computing perspective, in which information systems facilitate social interactions, informs the methods and techniques the approach uses. The objective is to design systems that respect social norms regarding information flows and address privacy in the context of collective information practices.13,14 This approach’s techniques can help design teams create interactions respectful of social and ethical norms. Feedback mechanisms about system functionality might help users evaluate the impact of system use on their privacy and change their future behavior accordingly. In addition to attending to individuals’ concerns, researchers evaluate the potential impact of complex information systems on groups of users and society in general. For instance, this approach attempts to answer ambitious questions such as whether we can develop mechanisms that use machine learning to reveal discrimination and social sorting, and what properties constitute a fair sociotechnical system.15

All three approaches fundamentally differ in what they consider to be privacy problems and solutions. This might be seen as productive plurality in research. However, isolation between the research communities; their varying positions on the role of technology, law, and society; and the distance between researchers and practitioners lead to gaps and vulnerabilities.

In practice, privacy-by-policy activities are often limited to privacy policy statements and checkboxes for consent and don’t result in changes to engineering practice or system design. These activities have mainly been the bailiwick of the legal team,10 members of which might not have an in-depth understanding of engineering privacy mechanisms’ potential and limitations. Purely technical approaches might prove insufficient for aligning nuanced legal policies with engineering artifacts and can fall short of addressing responsibilities across organizations. In the absence of normative guiding principles and evaluation, privacy-by-policy approaches might result in a set of procedures that fulfill compliance requirements but provide little effective protection. Moreover, top-down decisions to introduce data protection mechanisms, if insensitive to the organization’s engineering culture, might be met with resistance. In general, transforming existing practices might be a precondition for engaging engineers who feel that privacy is an abstract problem, not an immediate problem, not their problem, or not a problem at all.8

In contrast, in privacy-by-architecture approaches, the conception and implementation of privacy-protecting measures are mainly under the purview of technical experts with in-depth knowledge of cryptographic and traffic analysis techniques. The objective is to develop privacy tools or mechanisms that offer formal guarantees—that is, fulfill quantifiable privacy properties such as anonymity. Developing PETs requires mastering sophisticated engineering skills mainly acquired through participation in the community of experts. Efforts to integrate PETs into system engineers’ toolboxes or into larger systems have been limited. The absence of methods to implement, integrate, and maintain PETs and the scant attention given to socialization of the tools might pose obstacles to taking them from the lab into the wild. Even when experts integrate PETs into systems, they can face backlash,16 especially if the proposed mechanisms introduce usability or performance tradeoffs or meet political resistance.4

Purely technical approaches might prove insufficient for aligning nuanced legal policies with engineering artifacts.
Vulnerabilities might arise as a result of treating the different approaches as if they’re solutions that can be applied independently of one another. For instance, whether a social network service’s photo-tagging feature is more acceptable when tags are made public before or after the data subject’s confirmation, and whether these tags should be revocable varies depending on each community’s data-sharing practices. Because their focus is on tag design, UX engineers might treat photo storage and accessibility to the service provider as irrelevant to their task. However, such separation of concerns assumes that those potential risks that arise due to the underlying system architecture are independent of the local tagging practices. As a consequence, users might feel empowered in negotiating their privacy in social settings, while becoming increasingly vulnerable to violations of privacy by powerful service providers. Similarly, system engineers might constrain information flows in a way that strongly complicates and limits user-facing design. Especially in the context of PETs, such matters could lead to usability problems that dampen system adoption.

Finally, even if all three approaches are applied in concert, some privacy concerns might fall out of scope. Illustrative of such shortcomings is the tendency of all three approaches to produce solutions that scale only to a single organization—a model that doesn’t reflect the way new services, such as software as a service, are provisioned or the way free software projects are organized. Similarly, the Internet and mobile communication networks are examples of global infrastructure that require a different lens. Design decisions applied to infrastructures can have grave implications for privacy protections that can be applied to technologies built on them. Since the Snowden revelations, efforts to apply privacy-by-architecture methods and techniques in digital infrastructure design have gained in prominence, for instance, considering data minimization to protect against TLS client fingerprinting. These efforts have shown that addressing privacy in the Internet’s underlying protocols, Web browsers, or GSM standards is slow, complex, and readily dominated by those with the greatest resources to influence the process. Such processes can be stalled easily if, for example, the parties paying the tradeoff costs for privacy protection aren’t reaping the benefits. Although engineering methodologies can’t solve these political conflicts, they might help improve the process of developing inclusive and effective privacy solutions for global infrastructures.

**Building Blocks**

In defining the field of privacy engineering, we lean on software engineering, the subfield of computer science concerned with all aspects of the production of information systems, including the conceptualization, design, maintenance, and removal from service. Owing to the complexity of privacy as a social and legal concept, we also borrow knowledge and know-how from privacy research and practice.

Responding to the methodological shortcomings we described, we follow Sjaak Brinkkemper’s lead on method engineering and define privacy engineering as the field of research and practice that designs, implements, adapts, and evaluates theories, methods, techniques, and tools to systematically capture and address privacy issues when developing sociotechnical systems. In this context

- **privacy-engineering methods** are approaches for systematically capturing and addressing privacy issues during information system development, management, and maintenance;
- **privacy-engineering techniques** are procedures, possibly with a prescribed language or notation, to accomplish privacy-engineering tasks or activities; and
- **privacy-engineering tools** are (automated) means that support privacy engineers during part of a privacy-engineering process.

The definition would benefit from some further elaboration. First of all, what justifies identifying an engineering activity as pertaining to privacy? And, how can we answer this question if we don’t settle on a definition of privacy? As Deirdre Mulligan expressed elegantly in her IWPE keynote, privacy-engineering work requires embracing the plurality, contestability, and contextability of privacy as a social, political, and legal concept.

A primary example illustrating privacy’s plurality is the work of Daniel Solove, who distinguishes between the right to be left alone, limited access, control, personhood, secrecy, and intimacy based on an extensive study of torts in the US legal system.

Contextuality is best described using the justificatory framework developed by Helen Nissenbaum, who argues that privacy isn’t about control over or confidentiality of information, but rather ensuring appropriate information flows respectful of social norms in a given context. For example, during a consultation, it’s appropriate for a patient to disclose health information to the doctor, but not vice versa.

Contestability refers to the availability of multiple concepts around which disputes exist that can’t be settled by an appeal to “empirical evidence, linguistic usage, or the canons of logic alone.” Contestability provides a language for conversing about privacy’s meaning, allowing it to be flexible enough to capture very different privacy issues in rapidly changing sociotechnical systems.
(plurality) introduced in different contexts with varying information norms (contextuality).

To preserve its contestability, we refrain from folding a specific conception of privacy into the definition of the field. However, we do assume that any defendable privacy methodology will draw on some normative theory of privacy, be it legal, social, or political. In the absence of such a normative compass, we lack the parameters against which to judge whether a method, technique, or tool attends to privacy. For example, Fateme Shirazi and her colleagues compared experiences elicitation and architectural techniques—into the design process. Marit Hansen and her colleagues discussed a technique that would help engineers reconcile tensions between potentially conflicting privacy goals, like unlinkability and transparency. And finally, Fateme Shirazi and her colleagues compared experimental techniques and tools that engineers can use to assess the performance of or attacks on the Tor network without violating user privacy.

Privacy engineering foresees the use of these methods, techniques, and tools in developing information systems. By information systems, we refer to not only the technical artifact but the greater sociotechnical system. In using this term, we recognize that any system exists only in interplay with a host of social, political, legal, and economic arrangements. We argue that those in the privacy-engineering field need to be cognizant of the greater material and social networks that the engineered artifacts exist in. They should also support sociotechnical design practices that aspire to develop efficient and effective approaches to privacy and that, in the process, help improve the lives of those affected by these systems.

Three papers at IWPE beautifully teased out the sociotechnical aspects of addressing privacy in infrastructures. Nick Doty, after providing an overview of the methods followed by the Internet Engineering Task Force and the World Wide Web Consortium, described some tools to incentivize and evaluate the way privacy issues are addressed during the standards-making process. Gina Fisk and her colleagues presented a method to integrate existing privacy-engineering best practices—including privacy requirements elicitation and architectural techniques—into the design process. Marit Hansen and her colleagues discussed a technique that would help engineers reconcile tensions between potentially conflicting privacy goals, like unlinkability and transparency. And finally, Fateme Shirazi and her colleagues compared experimental techniques and tools that engineers can use to assess the performance of or attacks on the Tor network without violating user privacy.

Privacy engineering foresees the use of these methods, techniques, and tools in developing information systems. By information systems, we refer to not only the technical artifact but the greater sociotechnical system. In using this term, we recognize that any system exists only in interplay with a host of social, political, legal, and economic arrangements. We argue that those in the privacy-engineering field need to be cognizant of the greater material and social networks that the engineered artifacts exist in. They should also support sociotechnical design practices that aspire to develop efficient and effective approaches to privacy and that, in the process, help improve the lives of those affected by these systems.

Three papers at IWPE beautifully teased out the sociotechnical aspects of addressing privacy in infrastructures. Nick Doty, after providing an overview of the methods followed by the Internet Engineering Task Force and the World Wide Web Consortium, described some tools to incentivize and evaluate the way privacy issues are addressed during the standards-making process. Gina Fisk and her colleagues presented a method to minimize privacy risks in cybersecurity data sharing that prompted a discussion on the appropriateness of making privacy claims in infrastructures built for national security and surveillance. Eve Maler introduced an Internet-scalable consent mechanism that might prove to be the "sweet spot" that attends to both technical and regulatory challenges in the context of Internet of Things. In addition, our second keynote speaker, Ian Oliver from Nokia Networks, illustrated how privacy engineers can leverage concepts and techniques from the safety-critical domain. He highlighted checklists, dataflow modeling, and organizational roles as aids to enabling good engineering culture.

The field's robustness depends as much on the development of methodologies as it does on their implementation, adaptation, and evaluation. Reports and case studies that expose the challenges of implementing privacy technologies are elemental to the generalization and systematization of privacy-engineering knowledge. In their IWPE paper on secure two-party computation, Henrik Ziegeldorf and his colleagues implemented and evaluated the performance of different protocols to help nonexpert developers pick the framework that fits their needs. In the process, the authors documented implementation challenges unique to each protocol. Rainer Hörbe and Walter Hötzendorfer developed an evaluation technique for federated identity management systems that also can aid engineers in translating normative privacy principles into architectures.

In addition to addressing gaps in research, our definition of privacy engineering is comprehensive enough to encapsulate recent efforts in developing standardized processes. Ann Cavoukian and her colleagues defined privacy as a nonfunctional requirement in the engineering process, whereas MITRE and the National Institute of Science and Technology characterized privacy engineering as a form of risk analysis. These definitions frame privacy narrowly and constrain the type of methodologies that can be used to those that are risk based. They're skewed toward privacy-by-policy approaches and rarely attend to privacy-by-interaction methodologies. Informed by the diversity of research and practice, our definition of privacy engineering provides a broader framework in which existing and future efforts can be cultivated.

Looking Ahead
Efforts to address privacy using technical means are still scattered and disconnected. Few of these efforts explicitly attend to generalizing and systematizing associated
engineering practices so as to be accessible to a wider community. Public and private organizations’ continuing negative track record of privacy blunders suggests both would benefit from the development of a privacy-engineering practice.

Privacy engineering responds to these gaps, and IWPE is a forum where community members can come together to actively engage in the nascent field of this new field. At its first successful iteration, IWPE participants responded to the field’s challenges, identified gaps, and agreed on three issues that require urgent attention.

First, we need to develop methodologies to address concerns of parties’ increased capacity to use machine learning to draw inferences from datasets. With advances in software as a service, big data infrastructures, and artificial intelligence, greater inferences can be made about individuals and user populations. These inferences can be used to profile users, organize future interactions, and drive a shift to data-centric software engineering practice. What methods, techniques, and tools address surveillance, discrimination, and accountability concerns attributed to such semantic power in sociotechnical systems?

Second, we must conduct empirical studies that reflect on different contextual challenges to applying privacy-engineering methods, techniques, and tools. Implicit assumptions about system architectures, labor, expertise, and organization type underlie methods, techniques, and tools. Empirical studies that explore how privacy issues are (or aren’t) currently addressed in different engineering contexts and that evaluate which methods, techniques, and tools are more appropriate in a given context are crucial to the field’s success.

Finally, we need metrics and analytics to evaluate the efficacy of privacy-engineering activities. Metrics can be used to indicate the number of privacy violations, track the number of a system’s fulfilled privacy requirements, choose privacy tools, or evaluate privacy and performance tradeoffs. In some cases, rather than quantification, analytical evaluation based on interdisciplinary methodologies might be more appropriate. Both approaches are hot topics of future research.

These are a subset of the exciting challenges at the core of privacy engineering. We welcome the growing community of privacy-engineering research and practice to join us in further shaping this field at the next IWPE, to be held 25–26 May 2016, in San Jose, California, co-located with the 37th IEEE Symposium on Security and Privacy. Further information on IWPE 2016 can be found at http://ieeecomputer.org/TC/SPW2016/IWPE.

Acknowledgments
We thank our Organizing and Program Committee, the authors and attendees who contributed to the fruitful discussions held during the workshop sessions, and the IEEE Symposium on Security and Privacy workshops organizers for giving us the chance to kick off the First International Workshop on Privacy Engineering. We’re especially indebted to Helen Nissenbaum, Carmela Troncoso, Jaap-Henk Hoepman, and Yod-Samuel Martin as well as the anonymous reviewers of this article for their comments. This work was completed with the generous support of the Information Law Institute at New York University, the Center for Information Technology and Policy at Princeton University, a grant from the Flemish Research Council, and the PRIPARE project funded by the EU’s 7th Framework Programme under grant agreement ICT-610613.

References


Seda Gürses is a postdoctoral research associate at Princeton University’s Center for Information Technology Policy and an FWO (Fonds Wetenschappelijk Onderzoek–Vlaanderen) fellow at COSIC, University of Leuven. She works on privacy and requirements engineering, privacy enhancing technologies, and surveillance. Gürses received a PhD in computer science at the University of Leuven, Belgium. Contact her at fgurses@princeton.edu.

Jose M. del Alamo is an associate professor in the Information and Communications Technology (ICT) Systems Engineering Department at the Universidad Politécnica de Madrid. His research focuses on personal data management issues, including privacy and identity management, in the context of software and systems engineering. Del Alamo received a PhD in ICT systems engineering from the Universidad Politécnica de Madrid in 2009. Contact him at jm.delalamo@upm.es.