CCC
Computing Community Consortium
Catalyst

Privacy by Design- State of Research and Practice

Workshop 1 Report

Executive Summary

For years, lawmakers, advocates and engineers have touted the potential benefits of Privacy by Design, of integrating privacy throughout the entire technical design process rather than after-the-fact. Nonetheless, we still struggle with how to practice Privacy by Design, whether it is how to conceptualize privacy, how to build privacy in the engineering process, how to present those privacy designs to users or how to incentivize practice of and compliance with Privacy by Design.

In order to identify a shared research vision to support these different facets of the practice of Privacy by Design, the Computing Community Consortium (CCC) is sponsoring a series of four workshops throughout 2015. The first workshop took place in February in Berkeley, California, focusing on concepts of privacy and how different groups think about and approach privacy in their receptive domains. A group of over 40 collaborators were in attendance, representing various parts of industry, academia, government and civil society: from health care to social networking to telecommunications, from philosophy to law to computer science, from national intelligence services to state pension authorities to consumer advocates.

Here we describe the outcomes of the first workshop, highlighting the key insights, questions, themes, disagreements, and further barriers to actionable progress.

Key insights

Privacy is an "essentially contested" concept. In different contexts of life, "privacy" can mean something different, apply to different types of information, serve different interests and purposes, and in its absence lead to different harms. For example, privacy for a hospital protecting patients' privacy requires both preventing unwanted intrusions in patients rooms and limiting the use and disclosure of patient records; protecting a child's privacy may require providing a parent to limit how a service provider can collect and use their child's information, and providing the parent with access to any information that child divulges; privacy expectations and concerns can be associated with places, activities, records, relationships, in varying permutations—context—making it difficult to identify portable rules; privacy's absence could lead to intrusive surveillance or it might lead to discrimination by unwanted data aggregation and analysis. Using a series of case studies of privacy complaints arising in different sectors, groups of participants analyzed: the utility of existing privacy frameworks such as the Fair Information Practice Principles (FIPPs); taxonomies of privacy harms; concepts of privacy, such as Contextual Integrity; and other approaches for identifying privacy concerns. The group struggled with the "essentially contested" concept of privacy, but found that different concepts and analytical tools are helpful in the identification and potential mitigation of privacy concerns.

In the United States, there are many sources of privacy law, which reflect different conceptualizations of privacy. For example, the "right to be let alone" and "limited access to self" concepts of privacy are reflected in the privacy torts, as well as statutes that set conditions around commercial advertising coming into the home. The "limited access to self" and "secrecy" concepts of privacy exist in the 4th Amendment. "Control over personal information" is a key operational concept of the FIPPs. Newer concepts of privacy, such as Contextual Integrity, while not yet fully reflected in law are growing in importance as regulators attempt to limit the consent burden placed on individuals.

Research in computer science has produced a large variety of solutions for privacy, which operate at different levels of use and reflect different notions of privacy. This includes mechanisms such as auditing algorithms, database anonymization, multi-party computation, and others. These broadly conceptualize privacy in one of three ways: privacy as confidentiality, privacy as control over data and information, and privacy as social practice constantly subject to renegotiation in and across social contexts. These differing conceptions of privacy have driven divergent tools and solutions. Computer science research communities also differ in whether they research privacy solutions and tools at the level of individuals and users, at the level of organizations, or at the level of computing infrastructure. A significant component of existing research focuses on the collection of data. Participants discussed ways computer science research could focus on the use of data, to support user control over data use with minimal burden. Several ways of operationalizing conceptions of privacy into computer science research and engineering were discussed including:

- Taking privacy values or principles, translating them into properties, and then building technical mechanisms based on those properties. For example, principles might include things like "individual control" or "focused collection"; related properties might be "usable controls" or "data confidentiality"; and appropriate mechanisms might be "tagging data" or "homomorphic encryption".
- Examine individual privacy laws to derive rights and obligations, and use that to derive engineering and technical requirements.
- Use a set of organizing principles or framework, such as the FIPPS, and transform those into engineering guidelines or principles.

The workshop also heard "reports from the field" on those who have implemented — or are struggling to improve — privacy programs in the wild: at large tech companies, Internet standard-setting bodies or government agencies. Often highlighted were disciplinary differences: both in the different ways that academics (such as lawyers or computer scientists) approach the concept of privacy and its practice and in the effective organization of multi-disciplinary teams or groups within companies. Key insights included:

Industry

- **Implementing cross-functional privacy teams within companies** to review privacy implications of products. These teams could include a variety of expertise, including technical, project management, communications, and law.
- **Engaging in multiple types of research to better understand privacy**, such as user experience testing, user research in the field, surveys, and analysis of statistical data collected and generated by products, and integrating these understandings into products.
- **Developing educational tools for end users** that could help users better understand the types of privacy options a product offers such as tutorials, walkthroughs, or reminders to check privacy settings.
- **Using the agile development process** is a double edge sword; changing parts of the system during development, sometimes in response to emerging privacy concerns, can end up undoing other privacy solutions.
- **Creating privacy resources within the organization.** These could include **an internal ethical hotline** or **designating an individual who can focus on privacy aspects of a project through its entire development process** from end-to-end.
- **Developing access and use-based controls for data to protect privacy.** Focusing on the controlling and auditing use of data rather than the collection of data to preserve privacy.

Government Agencies

- **Using mathematical tools to protect privacy**. A desire to use mathematical quantification to get an indicator of risks when making privacy sensitive decisions with data, particularly when choosing among privacy engineering solutions such as homomorphic encryption or differential privacy.

- **Implementing technical standards for the protection of information.** Such as standards that allow the implementation of public-facing Federal programs that rely on users' data, but would not allow an agency to see any users' personal information.
- **Setting controls on use of data through internal standards and audits for privacy.** For an agency that does deal with personal information, in addition to technical access controls, it can create internal protocols and procedures on how employees can access and use data and what to do in case of a data breach. Foster a culture of privacy, to bolster intuitive understandings of what privacy requires in day-to-day practice.

Standards setting bodies have begun engaging more with privacy. They are exploring different ways to identify and address privacy concerns. Some standards setting organizations have tried to inculcate thinking about privacy into their processes by developing standards on privacy considerations, forming privacy review committees, and organizing programs to identify approaches for addressing privacy in internet protocols. Privacy can be the aim of a standard, however more often it arises as an ancillary property—or its loss as a side effect—of a standard with no explicit privacy goal. The National Institute of Standards and Technology (NIST) has taken a harm-based perspective on privacy, identifying a range of privacy harms and developing a risk management approach, with specific controls, to address them.

Engaging academics and practitioners from multiple disciplines and sectors is essential to develop a privacy research strategy that addresses the complexity of privacy in practice. Broad involvement is necessary given the different concepts of privacy, expectations, and related harms that manifest in various relationships and contexts. Social structures, communities, activities, technical artifacts, and pre-existing legal rights and obligations all shape privacy norms and experiences of privacy violations. For example, understanding privacy in a smart home environment versus privacy in a health care database differs because of differences in users, business models, expected use cases, user behaviors, technologies, applicable laws, potential harms, and social norms. Understanding this complexity requires situated analysis, diverse methods, and skilled analysis. Ideally, technical, social, ethical, and legal research is a coordinated or integrated into a synthetic understanding of privacy needs and problems in practice.

Key questions

From the workshop, several key questions were raised:

1. **How can we get tools and insights from computer science privacy research more widely adopted?** While a plethora of privacy solutions are studied in computer science, few have been widely adopted. What are the barriers to adoption, and how can these results be more widely adopted?
2. **How can privacy be operationalized in engineering and design?** Work with contested concepts, like privacy, that can vary across situations is challenging. What tools can assist engineers and designers in this work? Even compliance can be difficult due to ambiguities in the law, and the incomplete match or mismatch between legal concepts of privacy and the privacy concerns of relevant populations.
3. **What are the drivers of implementing privacy in the wild?** Is it "what the customer wants", legal compliance, best business practices, or other factors?
4. **How long should data be kept?** Data minimization was mentioned several times as one way to enhance privacy, but knowing what data to minimize and what data to keep, and how long to keep it can be difficult.
5. **What is the life cycle of data, with respect to privacy?** Is it just having the actual data or the entire process, from mode of collection through point of destruction?
6. **Who is responsible for privacy?** Who is responsible for privacy within an organization, who is responsible for what parts of privacy, and when are they responsible for it? How does privacy responsibility migrate over the full development of a project? Who should lead—a legal team, the engineers, product managers, or others? **Should hand-offs occur over the course of the project, and if so what tools are needed to create**

continuity across these actors given their different training and skills? What training and knowledge should the responsible parties bring to the table? What research can assist in answering this question

7. **What new tools are needed to bridge privacy conversations and implementations across the legal, ethical, and technical domains?** Is it best to identify properties related to various concepts of privacy that can be components of a system (some legal, some practice, some technical); or risks related to privacy failures that can be mitigated through the a similar systems approach; or some combination of the two? Relatedly, how much of privacy can be addressed in a fixed way, how much must be left to settle or configure later depending upon specific context or even during a particular event or interaction? How do we appreciate the contextual nature of privacy in practice—and leave sufficient flexibility to support it—while not burdening users? In particular, how should the limited knowledge of risks and benefits, and time devoted to privacy where it is an ancillary issue, shape implementation choices?
8. **What are the fields that impact privacy by design?** Privacy is clearly an area that is multidiscipline and needs to be explored as an interdisciplinary research area.

Disagreements

While there was largely agreement that the FIPPs do not cover all of today's privacy concerns, there was disagreement about the usefulness of the FIPPs moving forward. Should they be replaced or integrated into privacy work? Despite their shortcomings, they are widely established and can be applied as a scaffolding tool for practitioners working with privacy on the ground. At the same time they reflect a particular concept of privacy, and attend to specific kinds of privacy related harms – focusing on notice and consent for an individual at the moment of data collection – which seem to be narrower than what is needed in practice.

How should we approach defining or conceptualizing privacy? There are top-down approaches of creating taxonomies and organizing concepts of privacy, and there are bottom-up approaches focusing on user concerns and contexts. Each brings different concepts or analytical tools that can help us identify and address privacy concerns, and these approaches may not be mutually exclusive.

Barriers to Actionable Progress

Bridging the gap between research and operationalization or implementation of privacy tools. We need to better understand the barriers to adoption and why previous privacy focused projects have failed. This could include why technical projects have failed, as well as bringing in other expertise such as economists or sociologists.

There is a challenge within some organizations of understanding privacy as not only a data requirement or system requirement, but also as a business requirement. Privacy is more than an engineering problem, but one that encompasses the entire system, including the business aspects. Privacy engineering becomes less useful if other parts of the organization are able to negate it.

It is unclear where privacy expertise should be located. From an organizational standpoint, what types of people should be responsible for privacy, and how should they be educated? Should engineers be trained as privacy experts? Should business leaders? Should cross-functional teams be implemented in organizations? Where will these people be trained in privacy? What should be included in a "privacy curriculum"?

There is a need to bring people together from a wide range of research areas, but that brings its own challenges. The structure and availability of research funding affects the ways multi-disciplinary research groups form and affects what type of research gets funded. Furthermore, different fields may have different approaches to and conceptualizations of privacy, as well as different research norms, languages, and goals. Tools that bridge disciplines, or ease conversations across disciplines are essential.

Future Workshops

The next workshop, the second in the series of four, will be in May in Atlanta, GA at Georgia Tech and will focus on privacy from the perspective of designers. At the end of summer, the third workshop will be in Pittsburgh, PA at Carnegie Mellon University and it will bring together engineers to discuss privacy research and practice. Finally, to wrap up the series, the fourth workshop will in Washington, DC at Georgetown University to discuss legal and organizational research necessary to catalyze Privacy by Design.

Report Writers

Nicholas	Doty	University of California, Berkeley
Ann	Drobnis	Computing Community Consortium
Deirdre	Mulligan	University of California, Berkeley
Richmond	Wong	University of California, Berkeley

Workshop Participants

Annie	Antón	Georgia Institute of Technology
Alvaro	Bedoya	Georgetown University
Mike	Berger	University of California, Berkeley
Travis	Breaux	Carnegie Mellon University
Justin	Brookman	Center for Democracy & Technology
Sean	Brooks	NIST
Alissa	Cooper	Cisco
Anupam	Datta	Carnegie Mellon University
John	Delong	National Security Agency
Nicholas	Doty	University of California, Berkeley
Ann	Drobnis	Computing Community Consortium
Ed	Felten	Princeton University
Ed	Fok	DoT
Jonathan	Fox	McAfee
Robert	Gellman	Consultant with non-profits
Ari	Gesher	Palantir
Jesse	Goldhammer	University of California, Berkeley
Nathan	Good	Good Research
Susan	Graham	University of California, Berkeley / Computing Community Consortium
Seda	Gurses	New York University
Joe	Hall	Center for Democracy and Technology
Peter	Harsha	Computing Research Association
Jaap-Henk	Hoepman	Radboud University Nijmegen
Jen	King	University of California, Berkeley
Colin	Koopman	University of Oregon
Keith	Marzullo	National Science Foundation
Sigurd	Meldal	SJSU
Mary	Morshed	CalPERS
Deirdre	Mulligan	University of California, Berkeley

Erik	Neuenschwander	Apple
Helen	Nissenbaum	New York University
Nicole	Ozer	ACLU
Ed	Palmieri	Facebook
Audrey	Plonk	Intel
Tal	Rabin	IBM / CCC
Aaron	Rieke	Yu & Robinson
Thomas	Roessler	Google
Ira	Rubinstein	New York University
Fred	Schneider	Cornell University
Elaine	Sedenberg	University of California, Berkeley
Peter	Swire	Georgia Tech
Aimee	Tabor	University of California, Berkeley
Michael	Tschantz	University of California, Berkeley
Tomas	Vagoun	NITRD
Tara	Whalen	Google
Jeannette	Wing	Microsoft
Richmond	Wong	University of California, Berkeley
Helen	Wright	Computing Community Consortium
Scott	Young	Kaiser

This material is based upon work supported by the National Science Foundation under Grant No. (1136993).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.