Privacy Engineering:

Proactively Embedding Privacy, by Design

www.privacybydesign.ca

January 2014

Ann Cavoukian, Ph.D.
Information and Privacy Commissioner
Ontario, Canada

Stuart Shapiro, Ph.D.
MITRE Corporation

R. Jason Cronk, Esq.
Enterprivacy Consulting Group

Information and Privacy Commissioner
Ontario, Canada

2 Bloor Street East
Suite 1400
Toronto, Ontario
M4W 1A8
Canada

416-326-3333
1-800-387-0073
Fax: 416-325-9195
TTY (Teletypewriter): 416-325-7539
Website: www.ipc.on.ca
Privacy by Design: www.privacybydesign.ca

Privacy Engineering:

Proactively Embedding Privacy, by Design

TABLE OF CONTENTS

I. Introduction

Information management is a booming profession. The collection, use and disclosure of personally identifiable information ("PII") by organizations of all types around the world have grown dramatically in the past decade, along with the value of PII and the need to manage it responsibly. The enduring confidence of individuals, businesses, and regulators in organizations' data handling practices is a function of their ability to express core privacy commitments and requirements, which also promote efficiencies, innovation, and competitive advantages. Privacy *is* indeed good for business.

In response, we have seen the emergence, rapid rise, and professionalization of the corporate privacy officer tasked with applying Fair Information Practice Principles (FIPPs) and other international privacy standards such as the *Privacy by Design* Framework.[1] *Privacy by Design* Foundational Principles serve as an overarching framework for inserting privacy and data protection early, effectively and credibly into information technologies, organizational processes, networked architectures and, indeed, entire systems of governance and oversight. *PbD* seeks to raise the bar for privacy by promoting enhanced accountability and user trust.

If *Privacy by Design* provides the "what" to do, then *privacy engineering* provides the "how" to do it. While the term *privacy engineering* has been around since at least 2001[2], only in the past few years has it come into common usage in the privacy professionals' community. In the last two years positions for privacy engineers have begun to be advertised, and in Fall of 2013, Carnegie Mellon introduced its new program, a one-year Masters of Computer Science - Privacy.[3] As Lorrie Faith Cranor describes the role, "[a] privacy engineer is someone who understands the engineering and the privacy sides and works out strategies that allow people to protect privacy without getting in the way of building cool things."[4]

This paper is by no means exhaustive. A full treatment of privacy engineering would be voluminous. It begins with an introduction as to what privacy engineering entails, an acknowledgement that privacy is not strictly a technical concept (i.e. requires multidisciplinary considerations), and a look into how a privacy engineer approaches risks and risk analysis. Next, the broad classes of mitigating controls are considered. Finally, we briefly examine trade-offs; not between privacy and functional requirements, but rather against other considerations (costs, performance, etc.), and between the privacy implications of differing systems implementations.

[1] See Bamberger, Kenneth A. and Mulligan, Deirdre K., New Governance, Chief Privacy Officers, and the Corporate Management of Information Privacy in the United States (November 25, 2011). *Law and Policy* (accepted for publication in 2011); UC Berkeley Public Law Research Paper No. 1701087. Available at SSRN: http://ssrn.com/abstract=1701087

[2] Feigenbaum, Freedman, Sander, Shostack "Privacy Engineering for Digital Rights Management Systems", ACM Workshop on Security and Privacy in Digital Rights Management 2001

[3] See www.cmu.edu/news/stories/archives/2012/october/oct15_privacymasters.html

[4] Eric Hyle, "Q&A Privacy engineers could hold the key" *Tribune Review* Dec 12, 2012 Available at http://triblive.com/opinion/qanda/3123176-74/privacy-engineering-program

II. From FIPPs to *PbD*

Informational privacy is often said to be in “crisis” today, as a consequence of many trends and factors:

- leapfrogging information and communications technology developments;
- the advent of social, cloud, big data, mobile and ambient computing;
- variable cultural norms;
- the growth of ubiquitous computing;
- a global patchwork of evolving privacy laws and regulations.

The advent of networked information and communication technologies has, in one generation, radically changed the rules for managing data. Current trends carry profound implications for privacy. The creation and dissemination of data is accelerating globally, and is being replicated and stored everywhere, resulting in “oceans of data.” We can no longer speak meaningfully of information destruction, as we once did with paper records, because digital bits and bytes have now attained near immortality in cyberspace, thwarting efforts to successfully remove them from “public” domains. The practical obscurity of personal information – the data protection of yesteryear – is disappearing as data becomes digitized, connected to the grid, and exploited in countless new ways. We’ve all but given up trying to inventory and classify information, and now rely more on advanced search techniques and automated tools to manage and “mine” data. The combined effect is that while information has become cheap to distribute, copy, and recombine – too cheap to meter – personal information has also become far more available and consequential, and at the same time, more challenging to control and protect.

The privacy solution requires a combination of data minimization techniques, credible safeguards, meaningful individual participation in data processing life cycles and robust accountability measures in place by organizations informed by an enhanced and enforceable set of universal privacy principles better suited to modern realities.

Fair Information Practice Principles (FIPPs) have served as universal privacy values and as a common general framework for translating privacy objectives to law, policy, and technology. Many variants of FIPPs exist and are in force around the world today, varying in length, detail, and force of application. Despite superficial differences, they all share common fundamentals.

Privacy by Design evolved from early efforts to express Fair Information Practice Principles directly in the design and operation of information and communications technologies, resulting in *Privacy-Enhancing Technologies* (PETs). Over time, the broader systems and processes in which PETs were embedded and operated were also considered.[5]

[5] For an extended treatment of *PbD* origins, see Ann Cavoukian, “*Privacy by Design*: Origins, Meaning, and Prospects for Assuring Privacy and Trust in the Information Era,” in *Privacy Protection Measures and Technologies in Business Organizations: Aspects and Standards*, George O.M. Yee, ed (Aptus Research Solutions Inc. and Carleton University, Canada), pp. 178–208.

The broadening context for evaluating privacy risks and applying data protection principles goes well beyond a narrow focus on information communication technologies (ICTs) to include the "soft" legal, policy, procedural, and other organizational controls and operating contexts in which PETs might be embedded. A holistic, integrative approach to assuring privacy required taking into account developments in other areas, including:

- evolving legal and regulatory requirements;
- evolving organizational structures;
- evolving computing and networked contexts; and
- evolving consumer expectations and tastes.

A more preventative, practical, evidence-based approach to assuring privacy was necessary in the first decade of the millennium. This approach encouraged clear privacy promises to be made and conveyed. It meant emphasizing practical, measurable, and immediate results, based on universally-agreed-upon privacy values, common frameworks for integrating the diverse interests at play in exploiting personal information, and defined benchmarks for assessing adherence.

Privacy by Design Foundational Principles build upon universal FIPPs in a way that updates and adapts them to modern information management needs and requirements. By emphasizing proactive leadership and goal-setting, systematic and verifiable implementation methods, and demonstrable positive-sum results, *Privacy by Design* principles can assure effective organizational privacy and security by:

- serving as a framework for domain-specific control objectives and best practices;
- reducing harms and other "unintended" consequences associated with personal information;
- strengthening internal accountability mechanisms;
- demonstrating effectiveness and credibility of data management practices;
- supporting regulatory and third party oversight efforts;
- earning the confidence and trust of clients, partners and the public; and
- promoting market-based innovation, creativity and competitiveness.

III. What is Privacy Engineering?

In essence, *privacy engineering* is the discipline of understanding how to include privacy as a non-functional requirement in systems engineering. While privacy may also appear as a functional requirement of a given system (such as the TOR anonymity system), for most systems, privacy is ancillary to the primary purpose of the system. It may be required for compliance purposes, for customer trust, for risk management, or for ethical concerns but, in theory, the base system usually functions without privacy baked in.

Integrating privacy requirements into the areas of the systems engineering life cycle (SELC) can help facilitate core and other business objectives. For some organizations, the primary motive for privacy engineering will be for regulatory compliance purposes or reducing organizational risk. Beyond that, organizations may need to protect their reputation or brand in the market or leverage privacy as a differentiator or competitive advantage. The success of products such as Snapchat[6] show that there is clear demand in the market for privacy-protective technologies, even if people's purchasing decisions do not always rationally reflect their desires.[7] Some organizations also choose to preserve privacy because it is the right thing do, a form of "conscious capitalism" for the information economy.[8]

The role of the privacy engineer is to first identify which mechanisms govern or should govern information, and secondly, ensure that the system (and the organization) abides by that governance. Engineering for privacy necessarily requires multidisciplinary knowledge. A privacy engineer must:

- be aware of the legal and compliance obligations of the organization which operates the system;
- be aware of the privacy standards or international principals at play;
- understand the cultural norms of the population in which the system runs;
- identify privacy risks and help define the system privacy requirements;
- familiarize himself/herself with the company's systems engineering life cycle;
- be familiar with: data architecture, human-computer interaction, user experience design, and computer programming; and
- acquaint oneself with not only general privacy controls, but specific technologies to implement those controls.

In principle, privacy engineering shifts control to the data subject, but in practice this amounts to a shifting of control *toward* the data subject. However, the data subject may not always be in the best position to fully assess the risks involved. Privacy engineers and other privacy professionals may be better equipped to understand and address general privacy risks and issues. Just as users rely on security engineers to ensure the adequacy of encryption key lengths, for example, data subjects will rely on privacy engineers to appropriately embed risk-based controls within systems and processes.

[6] Cadie Thompson, "Why is Snapchat worth $800 million" CNBC, Jul, 24, 2013 Available at http://www.cnbc.com/id/100911089 The recent exposure of security flaws at Snapchat show that while consumers have pent up demand for privacy, companies need to do better to match their policies and technical capabilities with consumer expectations. Failure to do so erodes consumer confidence and trust.

[7] Allesandro Acquisti and Jens Grossklags, "Losses, Gains, and Hyperbolic Discounting: An Experimental Approach to Information Security Attitudes and Behavior" May 2003 2nd Annual Workshop on "Economics and Information Security."

[8] John Mackey, Conscious Capitalism: Liberating the Heroic Spirit of Business 2013 Harvard Business School Publishing

The common misperception is that information security equates to privacy. While security certainly plays a vital role in enhancing privacy, there is an important distinction to be made. From an organizational viewpoint, security is about protecting and controlling information. Encryption, identity and access management, firewalls, etc. are all about controlling the access or flow of information within the organization or between the organization and outside entities (be they vendors, customers or others). Privacy is the other side of the coin. It is about recognizing that the dominion of the information no longer rests with the organization.[9] Though they may retain physical control of data, the decisions about personal information are governed by law or regulation, by cultural norms, by unilateral policy, by contract, by economics, or by technical controls that implement individual consent and personal preferences.[10] Security is used to enforce those decisions, but not to make the decisions.

Security engineering is a relatively mature field with significant research behind it. Privacy engineering is more nascent and more subjective, with the term's definition not generally agreed upon in the industry. Security engineering coaches designers to understand the vulnerabilities of a particular system design and attempts to mitigate those vulnerabilities. While some non-technical considerations need to be weighed, and decisions made between various control strategies in security engineering, privacy engineering requires much more of a balancing of policy objectives.[11]

IV. Non-technical Considerations of Privacy

Privacy is socio-technical.[12] In other words, there are both socio-cultural and technical aspects to privacy. Focusing solely on the technical aspects of privacy in systems engineering invites failure. A system must be user-centric; user-centricity requires attention to both technical and non-technical considerations, including the user's thoughts and behaviors and how the user fits within the broader society.

User-centric Design

The tension between usability and privacy appears to be a constant battle. It most often arises in how best to engage the user in FIPPs-based notice and choice regimes but it can also arise in the context of a privacy-preserving system

[9] Casey Johnston, "Snapchat's bad security show how data use policies fail" *ArsTechnica,* Jan 6, 2014 Available at http://arstechnica.com/tech-policy/2014/01/snapchats-bad-security-shows-how-data-use-policies-fail/

[10] An important side note. While this paper primarily talks in terms of information privacy, the basic concepts apply to other forms of privacy (namely physical and decisional).

[11] Privacy engineers need to work closely with legal / compliance team. See Peter Swire and Annie Antón, "Engineers and Lawyers in Privacy Protection: Can We All Just Get Along?" in IAPP *Privacy Perspectives,* January 13, 2014. Available at http://bit.ly/1aooBBD

[12] Seda Guerses, Camela Troncoso, and Claudia Diaz, "Engineering for Privacy by Design"

that is not based on notice and choice.[13] Tension most often occurs during the design decision making process. Notice and choice can be interruptive of the user experience and detract from the functional usability of the product or service. But burying notification of privacy risks and the choice mechanism outside the immediate attention of the user fails to adequately put the user on notice, defeating the purpose.

Quality user experience design uses contextual clues to put the user in an analogous situation where he/she can appreciate the risks and potentially take action to mitigate those risks. Even the best designs may fail because new technology makes use of information in ways that may be unexpected, not conforming to norms with which we have become familiar. The privacy engineer must balance those privacy risks that may be presented to the user for decision-making, and those that are best handled through other means.

The user experience should flow naturally from a default of privacy. The user interface should suggest the natural collection and use of information to perform whatever service the user is expecting. Extraordinary collection and use which deviates from the user's expectation should require proactive action to be taken on the part of that user. Wherever practicable, users should be able to exercise meaningful information control options.[14]

Mutuality of Expectations

Most privacy expectations are based on un-codified norms which have evolved socially and culturally. Humans have developed these expectations to fit a variety of the social interactions we engage in. When we share information in certain settings (doctor's office, a friend's house, a marriage), there exist unstated ground rules about how the information should be used and disseminated. If we are concerned that the existing cultural norms are insufficient, an information provider may use additional contextual clues to aid the recipient in crafting the mutual understanding: a friend may whisper in another's ear, a customer may write his account number down rather than announcing it, etc. While these actions may be privacy preserving in their own right, by preventing inadvertent disclosure to nearby persons, they also provide a non-verbal hint to the recipient that he/she should take due care in who he/she shares that information with. If those clues are insufficient, people have another method by verbally communicating additional restrictions. "Can you keep a secret?" is a common phrase meant to impart a clear need for confidentiality.

Education, awareness, visibility and transparency all help to shape mutual expectations. It is important, not only that the privacy engineer understands the expectations of others, but helps to shape those expectations, especially in situations without clear historical norms. Communicating the intended uses or future dissemination plans is imperative in order to adhere to the *PbD* Foundational

[13] Jeremy Clark, P.C. van Oorschot, Carlisle Adams, "Usability of Anonymous Web Browsing: An Examination of Tor Interfaces and Deployability" Symposium On Usable Privacy and Security (SOUPS) 2007

[14] Traditional "notice and choice" regimes are giving way to more robust "transparency and control" mechanisms. For a discussion in the online context, see Ann Cavoukian, Ph.D. and Justin B. Weiss, J.D., 2012. *Privacy by Design and User Interfaces: Emerging Design Criteria – Keep it User-Centric*, Accessed at: www.ipc.on.ca/images/Resources/pbd-user-interfaces_Yahoo.pdf

Principle of visibility and transparency. This is not only necessary for users, but other stakeholders who will help shape the fluid cultural norms.

Behavioral Economics and Human Irrationality

Mutuality of expectations exists when our norms are clearly developed. Unfortunately, technology creates interactions in which no existing norm applies, either directly or by analogy. This creates a disjuncture between our privacy preferences and our actions or those of others. Allesandro Acquisti has shown repeatedly that we don't necessarily follow our own stated positions when making decisions affecting our privacy[15]. The decisional calculations are too difficult for us to make when they don't fit our existing mental models. The addition of cognitive biases, such as hyperbolic discounting[16], creates additional hurdles that a privacy engineer must overcome. Being user-centric means more than just giving the user choice in the use and disclosure of his/her information. The privacy engineer must ask whether the users' purported choice is truly representative of their privacy preferences or designed to take advantage of their cognitive limitations. Disclosure should not only be voluntary but that voluntariness should be unhinged by the subject's inability to fully incorporate his/her preferences into his/her decision-making process.

Proportionality

Particularly important in privacy engineering is the need to consider the proportionality of the proposed system or solution within a societal context.[17] Does the proposed system cause more risk of harm to the affected group than the derived benefit? While similar to the use limitation principle of FIPPs, "proportionality … differs in one important aspect …, in that it establishes a balance between the usefulness of the considered application and its effects on privacy, whereas FIPPs only impose conditions on the collection and use of the information in relation to the data subject's consent, and to the needs of the application."[18] Similar to the behavioral economics concern identified above, the pendulum is swung from a focus on respecting the user's stated position to respecting the user, in toto.

Iachello and Abowd divide proportionality into three stages: legitimacy, appropriateness and adequacy. Legitimacy speaks to the question of whether the purpose of the system outweighs the potential downsides. Once legitimacy is established, the appropriateness of various implementing technologies needs to be considered. Finally, does the chosen technological implementation adequately mitigate the known risks? This goes to the heart of the specific design features of the system. They provide, as an example, a doorbell camera that is constrained to only activate and capture persons at the entrance of the door while avoiding

[15] See footnote 7

[16] Hyperbolic discounting, also called present bias, is the tendency to discount or ignore future benefit (or harm) against present benefit more than rationally reasonable.

[17] Giovanni Iachello and Gregory Abowd, "Privacy and Proportionality: Adapting Legal Evaluation Techniques to Inform Design In Ubiquitous Computing", CHI 205

[18] Ibid.

capturing the surrounding neighborhood. In this case, the narrowing of the camera viewpoint appropriately augments the camera functionality for privacy by mitigating the risk that people in the street may be inadvertently recorded.[19]

Proportionality arguments are frequent in the biometric space. The Privacy Commissioner of Hong Kong ordered a school to stop collecting fingerprints of school children to track attendance not only because the children couldn't appreciate the nature of what they were giving away, but because the use to which it was put lacked any sense of proportionality.[20] In other words, there were other, less privacy-invasive, ways to achieve the same objective.

V. Risk Models

Risk is a key aspect of systems engineering. Privacy engineering, therefore, requires a sufficiently robust approach to risk. This includes characterizing systems and processes in a way that is amenable to analysis using an appropriate risk model. Identified risks are then addressed through risk management approaches involving the selection and application of risk controls. Characterization tends to be relatively straightforward and can usually be carried out in terms of flows and changes in the states of personal information. Risk management also tends to be straightforward, at least conceptually, with standard control options including mitigation and acceptance. Defining and using an effective privacy risk model, on the other hand, is more involved.

Strictly speaking, a model based on compliance with privacy-related statutes, regulations, and policies is the most common privacy risk model, though it may not be explicitly viewed that way. However, given that most of those statutes, regulations, and policies are based on FIPPs, in general FIPPs serve as the most common substantive privacy risk model. A risk model of any kind establishes conceptual scaffolding to support structured reasoning about risks in a particular domain as represented by threats, vulnerabilities, and impacts. Privacy is one such domain and analytical frameworks based upon traditional FIPPs represent one such risk model. However, the increasing novelty and complexity of modern socio-technical systems and processes presents challenges for this model.

Fortunately, recent research has produced bases for additional privacy risk models that can help to identify those risks that a FIPPs-based risk model may fail to detect by itself. These works don't constitute complete privacy risk models in and of themselves, but they do provide grounded reference points around which more complete models may be built. One of the best known of these is Solove's taxonomy of privacy problems[21], which categorizes privacy harms into 16 categories comprising the four more general groupings of information collection, information processing, information dissemination, and invasions. These individual types of privacy harm equate to impact from a risk model standpoint. With a set array

[19] Ibid.

[20] Mari Shroff "Protecting Biometric Data: Privacy By Design", *2010 Biometric Institute of New Zealand Conference*, (Mar 26, 2010)

[21] Daniel Solove, *Understanding Privacy*. Cambridge: Harvard University Press, 2010.

of impacts serving as potential endpoint risks, analysis can work backwards to establish whether threats and vulnerabilities exist that could result in those risks.

Also well known is Nissenbaum's contextual integrity heuristic[22], which posits that privacy problems result from disruptions to informational norms related to what personal information flows from, and to which actors, under what conditions. Rather than provide a set of impacts, contextual integrity supports the identification of vulnerabilities in the form of disruptions that could result in privacy risks. In this case the reference point is located in the middle and requires working backward to identify threats that could exploit the vulnerabilities, and forward to the impacts that could result.

Both Solove's and Nissenbaum's approaches are essentially bottom-up. Rather than establish overarching principles for identifying privacy problems, they seek to capture those issues that become associated with privacy – *in situ*, in the case of contextual integrity. This contrasts with both FIPPs and with Calo's objective/subjective framing of privacy harms. Under Calo's scheme[23], subjective privacy harm results from an individual's perception that he or she is under observation, broadly construed, irrespective of the veracity of that perception. Objective privacy harm, on the other hand, results from the actuality of coerced disclosure of personal information or its unanticipated use. Both types represent impacts, but objective harms are external to the individual while subjective harms are internal. As with Solove's taxonomy, one must work backward to identify threats and vulnerabilities that could produce these risks. However, Calo's risk types are much broader than Solove's and, depending on one's proclivities, might make the job of constructing a larger risk model easier or more difficult. In any event, Calo's approach provides another reference point for privacy impacts that can then be linked to applicable vulnerabilities and threats.

All of these models are arguably more user-centric than a compliance-based risk model and therefore serve to complement a compliance-based risk model in a vital way. By focusing on individual and societal vulnerabilities and impacts rather than purely on potential violations of regulatory obligations, privacy engineering shifts the focus back to the user writ both large and small.

VI. Risk Analysis

Once an appropriate risk model (or model fragment) is identified, there are several considerations for performing the actual analysis. The analysis needs to be done at an appropriate juncture in the system engineering development life cycle (SELC) to be meaningful and provide sufficient time to mitigate those risks. The privacy engineer must examine the organization's SELC to determine the appropriate placement but ideally, it should be early in the process.

[22] Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Palo Alto: Stanford Law Books, 2009.

[23] M. Ryan Calo, "The Boundaries of Privacy Harm," *Indiana Law Journal,* Vol. 86, No. 3, 2011.

Checklists and Privacy Impact Assessments

The checklist approach to privacy protection has been debated.[24] Checklists have become important safety elements in airplane and medical procedures and are quite common in security auditing. However, their utility for privacy remains questionable. It might be possible to design privacy checklists for frequent and standardized use cases, but the breadth of potential projects makes a standard checklist for everything an unlikely tool.

A privacy impact assessment (PIA) can be a tool to support a formal risk analysis.[25] If the PIA uses a robust risk model, such as those mentioned above, and is augmented, where feasible, with metrics to determine the likelihood of the privacy impacts, the PIA becomes an important tool for performing the risk analysis. The PIA should not be a tool for analyzing business risks per se, i.e. the risk of regulatory or civil liability. The PIA should be user centric and analyze the risks to the data subjects. Each risk can then be reviewed in light of the regulatory environment, the competitive advantage and the corporate ethos, to determine whether mitigating controls are warranted.

Quantifying Privacy

Risk is a function of harm and probability. Quantifying privacy harm, though, is problematic as it is extremely dependent on the norms of the subject population as well as the availability of data upon which to base probabilities. Revealing a teen's pregnancy to her father may be thoughtless and embarrassing in one culture but deadly in another. While privacy violations which result in economic loss (such as stolen credit card numbers) are easier to quantify, most privacy violations are inherently more subjective and diversified in impact across subject populations.

In large scale systems, applying quantitative metrics to privacy compliance can be an important tool in ongoing analysis and risk mitigation. Proactive monitoring of known violations as they occur and quantitative reports can help justify expenditure on more robust controls or provide evidence that such expenditure is not warranted.

Residual Risks

Risk, be it privacy or any other kind, can seldom be entirely eliminated (except by avoiding the relevant activity or practice altogether). Regardless of which controls are employed, some type and degree of risk likely will remain. Even after privacy risk controls are applied, the residual or remaining risk must be

[24] Ian Oliver, "Safety Systems – Defining Moments" Available at http://ijosblog.blogspot.com/2013/07/systems-safety-defining-moments.html

[25] For a thorough review of the state of the art of PIAs, see David Wright, Kush Wadhwa, Paul De Hert, & Dariusz Kloza, Eds. *A Privacy Impact Assessment Framework for data protection and privacy rights*, Sept 2011, JLS/2009-2010/DAP/AG at: www.piafproject.eu/ref/PIAF_D1_21_Sept_2011.pdf

assessed. Thus, privacy risk assessment must be a continuous process rather than a one-time exercise.

This does not entail completely reworking risk analyses from scratch, likely an untenable option in any case when dealing with complex systems and processes. Rather, it involves reworking the specific applicable elements of the risk assessment starting with the relevant characterizations and moving through the affected analysis and controls. Upon reaching the risk control phase, one judges whether the residual risk reflected in the analysis (resulting from the application of the previously selected controls) requires further measures.[26]

> Minimizing some risks may create new ones, or exacerbate others. The choice of system architectures (e.g. centralized or distributed) can have profoundly different threat and risk scenarios. Detailed logging of network activities, or creating a registry to identify and prevent unauthorized behaviour, creates new risks if the registry or logs are not secure. Strong security controls established at one point can sometimes induce staff to adopt more convenient, but riskier workarounds. Policies and procedures to disclose breach details to customers via postal mail create new risks if customers have moved without a forwarding address.

Trade-offs

Systems engineering is typically replete with trade-offs – privacy engineering is no exception. Typically, trade-offs will include the classic one of schedule versus cost. However, this is only the most obvious and well-known trade-off. Others may include reliability, security, usability, performance, and functionality of the resulting system or process. This mix may include privacy as well, even while maintaining adherence to the positive-sum principle of *Privacy by Design*. Two distinct reasons underlie this.

First, the positive-sum principle deals with functionality, arguing against a *prima facie* assumption that privacy must be sacrificed to achieve some specific desired functionality. Such assumptions are generally unwarranted and a thorough exploration of the relevant trade space will confirm the often fallacious nature of such assumptions. However, this does not rule out the possibility that some other type of trade-off consisting of privacy versus something other than functionality – performance or cost, for example – may prove salient. Indeed, it has often been observed that the essence of engineering is satisfying goals or needs within constraints. Secure multi-party computation, for example, currently suffers from high computational overhead. One alternative is to mimic secure multi-party computation through clever data manipulation. Such approaches, though, do not offer the same mathematical guarantees of privacy. Therefore, the applicable trade space will involve trading off mathematical assurance of privacy for performance and vice versa.

[26] See Office of the Information & Privacy Commissioner of Ontario, Ontario Lottery and Gaming, and YMCA of Greater Toronto. 2010. *Privacy Risk Management: Building privacy protection into a Risk Management Framework to ensure that privacy risks are managed, by default*, and Cavoukian, Ann & McQuay, Terry, 2009. *A Pragmatic Approach to Privacy Risk Optimization: Privacy by Design for Business Practices*

Second, situations may also arise in which one aspect of privacy must be traded off against some other aspect of privacy. For example, imagine designing a system that requires the use of a biometric. Fingerprints are one of the most common biometrics currently in use and as a result, there exist significant sources of auxiliary knowledge that can link this biometric to biographical information, such as a name. Retinal biometrics, on the other hand, are less commonly used, with a commensurate reduction in the availability of auxiliary knowledge. However, retinal scans also have the capacity to reveal information regarding the health status of individuals as retinas are involved in the pathologies of certain diseases, including diabetes. Similar trade-offs exist for other biometrics. There is a trade-off, therefore, between the potential linkability of the biometric to other information and the information that may be revealed by the biometric itself. While controls potentially could be deployed to mitigate the risk incurred by a specific trade-off, this does not alter the fact of the trade space's existence. Rather, this highlights the importance of recognizing such trade-offs so that design and mitigation decisions can be made with full cognizance of their ramifications. Of course, one might also consider using a non-biometric mechanism instead, potentially resulting in a trade-off more akin to the type previously discussed, such as privacy versus usability.

VII. Controls

There is no clear consensus on a taxonomy of control strategies for enhancing privacy in systems. Spiekerman and Cranor, in their work on Engineering Privacy, identify a dichotomous taxonomy: privacy by policy and privacy by architecture.[27] Privacy by architecture focuses on reducing identifiability of data and of network centricity. Privacy by policy involves a "trust us" mentality to do the right thing for users, while privacy by architecture involves "trusting the system." A third, hybrid approach, not suggested in the Spiekerman and Cranor paper involves technical point controls to increase information privacy of specific system aspects.

Privacy by Policy

Privacy by policy is the most common approach taken by organizations today. Most of the policies follow legal or regulatory compliance requirements in the jurisdiction in which the organization operates. Some policies exceed minimum legal requirements and if published in a privacy statement, those policies may become subject to enforcement though civil actions by individuals or regulatory bodies (such as the FTC in the United States or Data Protection Authorities in Europe).

Policy is a way of managing the organization so that it interacts in a unified manner with others (be they users, consumers or other stakeholders). Policy must be a guiding force when implementing *Privacy by Design*.

[27] Sarah Spiekerman and Lorrie Faith Cranor, "Engineering Privacy", *IEEE Transaction on Software Engineering*, Vol. 35, No. 1, January/February 2009

Simply because a policy is an abstract concept does not mean that there can't be an engineering component. Privacy engineers play a critical role in shaping and implementing policy and how the organization implements those policies:

1. Identify cultural and social norms – Policies should not just be a function of compliance obligations but should be guided by the context in which the organization and the system sit. The privacy engineer will help identify this context and seek to understand and incorporate the expectations of users into the system or promote alterations to the context and thus expectations.

2. Inform policy-makers of the available technical controls – The adage "when your only tool is a hammer…." applies here. Policy-makers must know the technical capabilities and limitations so the policy controls can be implemented where architectural or technical controls are unavailable or unwarranted. Privacy engineers provide the necessary guidance for the available tools.

3. Develop policies for the SELC – Addressing privacy risks must be embedded into the system development life cycle. The privacy engineer should develop policies governing the SELC and those involved in the process so that privacy concerns are effective and not overlooked. This could be considered a meta-policy which requires policy considerations be examined during system development. The Google Streetview WiFi case provides a good example in this regards.[28] While there may have been a policy not to collect extraneous data or not to collect data without proper prior review, it failed because the processes weren't in place in the systems engineering life cycle to implement and enforce such a policy.[29]

4. Provide training input on proper and improper system use – Regardless of the technical controls, those with legitimate access to a system must understand how to use the system properly and not inadvertently incur privacy violations. The privacy engineer can provide guidance on developing proper training methods that combine requisite policies with actual demonstrations of system use.

Privacy by Architecture

Far less common in current practice, privacy by architecture seeks to ensure data subject privacy through the architecture of the information system by removing unnecessary data elements or pushing necessary data elements from the sphere of organizational control to the data subject's control. Spiekerman and Cranor describe this as anonymization and decentralization. Privacy by architecture provides a proscriptive recipe for win-win, full functionality that *Privacy by Design* requires. The goal is to provide a solution that performs the business function that the system was built for and does so in a privacy preserving manner. A few other architectural strategies, such as data minimization and PETs (privacy-enhancing technologies), are also described below.

[28] Edward Wyatt, "Denials over Google Streetview" *NY TIMES*, June 22, 2012 at www.nytimes.com/2012/06/13/technology/google-street-view-case-brought-employee-denials.html

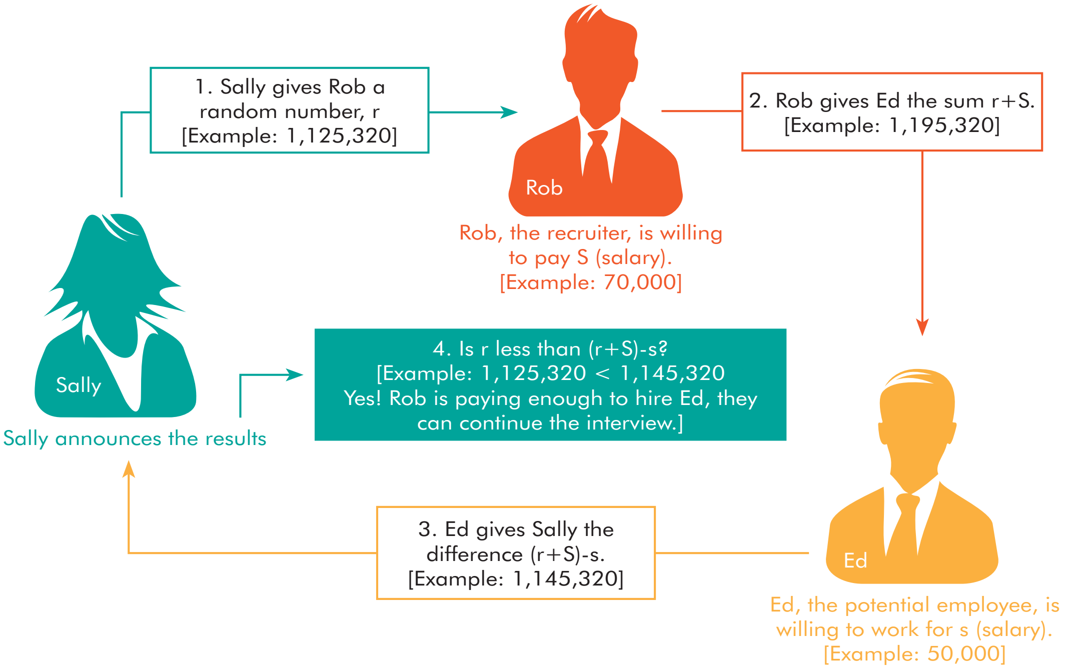
[29] See Cavoukian, Ann and Cameron, Kim. 2011. *Wi-Fi Positioning Systems: Beware of Unintended Consequences* at: www.ipc.on.ca/images/Resources/wi-fi.pdf

i. Data minimization

While data minimization can be a policy ("Don't collect unnecessary information") or a technical control (using structured entry forms to prevent extraneous data entry), it can also be an architectural strategy. The distinction exists where information seemingly essential to the business function is not collected by the organization itself, but rather the processing is done in such a way that the organization never possesses the data. This is best illustrated by example. Suppose a recruiter, Rob, is calling a potential employee, Ed, about setting up a job interview. One of the first questions Rob wants to know is whether Ed is willing to work for the salary the company is paying. Ed, not wanting to give up his negotiating position at such an early stage, does not want to reveal his true salary requirements.

Sally, knowing about the conundrum faced by Ed and Rob, decides to provide a service to them comparing Ed's salary requirements to the company's pay while only revealing the results to the parties. A simple approach to the problem would be for Ed and Rob to whisper into Sally's ear their respective amounts and have her do a quick comparison to provide the answer. However, there is still a privacy risk: Sally could be judgmental of Ed ("oh, he only earns THAT much?"); she could sell the information on what Rob's company is paying its employees to his competitor. Ed and Rob are reluctant to reveal this information to Sally. Sally must do more than the "trust me" privacy by policy approach to win their business.

Sally decides to architect her system in a way to minimize the data she needs. She doesn't need the actual salaries. What she really only needs to know is which figure is larger. She decides instead to hand Rob a random number. Rob, in turn, adds the amount his company is willing to pay to the random number and gives the result to Ed. Ed subtracts his salary requirement from that number and provides the result to Sally. Now Sally compares the original random number and the amount Ed provided her. If the original number is higher, the company is willing to pay Ed more than he is willing to work for. On the other hand, if the original number is lower, then Ed's salary demands are too high for Rob's company.



1. Sally gives Rob a random number, r [Example: 1,125,320]
Rob
Rob, the recruiter, is willing to pay S (salary). [Example: 70,000]
2. Rob gives Ed the sum r+S. [Example: 1,195,320]
Sally
Sally announces the results
4. Is r less than (r+S)-s? [Example: 1,125,320 < 1,145,320 Yes! Rob is paying enough to hire Ed, they can continue the interview.]
3. Ed gives Sally the difference (r+S)-s. [Example: 1,145,320]
Ed
Ed, the potential employee, is willing to work for s (salary). [Example: 50,000]

Data minimization has changed the architecture of the system from one in which Sally knows everything to one in which no one learns the direct information of any other participant but does learn the answer to the overriding question of whether Rob is paying a salary higher than Ed's requirement.[30] Further, she has mitigated the risks to Ed and Rob identified above. While the first approach focuses on Ed and Rob's trust of Sally to do the right thing, this second approach requires far less trust in the participants and more trust in the process.[31]

The design above involves anonymization and decentralization, which are described further below, but for the purpose of data minimization on the part of Sally. The solution further anonymizes the process because Sally no longer need keep track of a recruiter and potential hire but rather a random number and the results. It also decentralizes the system by having the clients perform some of the data processing.[32] Such a design is emblematic of the positive-sum *PbD* principle whereby both system functionality and privacy are preserved – win/win!

ii Anonymization

Anonymization is a specialized class of data minimization focused on the explicit avoidance, separation, or removal of the identifying information of data subjects from information systems. Many organizations' decision to identify their customers or data subjects stems not from a need but rather a common desire to know who they're dealing with. Collecting identifying information, though, in many cases may not be necessary to perform the primary function of the system, but the architects justify it based on a secondary or supporting function (for instance payment collection, fraud prevention, or even improving the civility of the user base[33]).

Anonymization is not just about the non-collection of personally identifying information. Identifiability can come from those data elements or patterns matching other data with auxiliary data sources. Pattern matching may come in many forms. Examples include the now infamous Netflix release of movie viewing habits which was matched against Internet Movie Database (IMDB) posts to identify some users based on their review history[34] or the potential use of a small set of mobile phone locations to identity people.[35]

"*Engineering Privacy*" divides anonymization into four stages of identifiability with increasing privacy protection.[36] Stage one is ***identified***, where unique

[30] The implication shouldn't be that Ed and Rob actually have to break out a pencil and do addition and subtraction; rather their client side applications could be developed to perform this function transparently to them.

[31] The requirement for trust in the participants and the resultant privacy risk are not completely lost in this design. Sally and Rob could collude to determine Ed's salary requirement and similarly Ed and Sally could collude to determine what Rob is willing to pay. While the impact of such a breach is the same in this case, the probability of collusion is presumably lower and thus the risk is lower.

[32] Another solution where the clients perform the entire operation is presented in the PETs section below.

[33] R. Jason Cronk, "FaceBook, Real Names and Social Circle Segmentation" http://privacymaverick.com/2013/03/14/facebook-and-real-names-and-social-circle-segmentation/

[34] Arvind Narayana and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets" available at v http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

[35] Jason Palmer, "Mobile location data 'present anonymity risk' BBC News Science and Environment, March 25, 2013 Available at http://www.bbc.co.uk/news/science-environment-21923360

[36] See Footnote 21

identifiers cross a database which also stores personal information. Stages two and three span the ***pseudonymous*** range, increasing degrees of difficultly in linkability. Finally, stage four, ***anonymous***, avoids not only unique identifiers or contact information, but the storage of data which could be combined with other data outside the system to achieve identification. As identifiability in the system decreases, there is a shift from "privacy by policy" and a need to dictate rules to govern the information to "privacy by architecture," whereby the lack of identifiable information limits its usability.[37]

iii. Decentralization

The other architectural concept explored by Spiekerman and Cranor is decentralization, the pushing out of data collection and processing to intelligent clients rather than processing in a centralized system. The privacy benefit to decentralization is that it pushes control over information out to the client (the data subject), generally, and it removes the economies of scale that an adversary has when breaching a centralized system. However, there could be a trade-off in security depending on the nature of the processing platforms. Mobile devices, for example, could be lost, and depending on configuration, could compromise privacy to those individuals.

iv. Privacy-enhancing Technologies

There is no clear cut definition of "privacy enhancing technologies" (PETs).[38] In fact, there is discordance between the business community and academic community, with the former more likely to label any technology that supports data subject privacy as a PET and the latter being much more circumspect in its application. However, there is a clear demarcation between privacy supportive technology (such as identity and access management tools) which may serve other goals, like maintaining data integrity and availability, and those technologies that exist **solely** to enhance privacy. A classic example of a PET is a mix network which serves to remove linkability in online communications. The sole purpose of the mix network is to enhance privacy, while still providing the underlying functionality (i.e., to facilitate communications).

PETs most often implicate the architectural design of an information system and thus must be considered early on. In returning to the example of the employee and recruiter, Ed and Rob, a PET solution would eliminate the need for Sally to do anything. Andrew Yao's millionaires' problem allows two millionaires to compare their respective wealth without either having to reveal the figures to anyone. Yao's solution uses a secure multi-party computation to determine the answer without revealing any information other than the results.[39] For instance, our computation of Ed salary requirements and Rob's offering would imply their value relative to each other, but nothing else.

[37] Statistical techniques and methods to de-identify data are possible, See, for example, the extensive work of Dr. Khaled El Emam and Commissioner Cavoukian, among others, available at: www.privacybydesign.ca/index.php/de-identification-centre/de-identification-tools-and-guidance/

[38] "Study on the economic benefits of privacy-enhancing technologies (PETs)" Final report to the European Commission DC Justice, Freedom, Security 2010

[39] Overview available at http://en.wikipedia.org/wiki/Yao%27s_Millionaires%27_Problem

Technical Point Controls

Privacy risks that aren't mitigated by the architectural strategy can still be reduced through additional technological means. Generally, these point controls seek to reduce the amount of information an organization has, obscure the meaning of the information or make it more difficult or costly to obtain the information.

i. Data Minimization

We've already talked about data minimization as a policy and as a guiding architectural principle. Data minimization can also be employed as a point control at the point of collection to avoid taking in unnecessary information. Machine to machine interfaces may employ data minimization techniques to reduce inadvertent disclosure from other parts of the system. Aggregation can also serve a data minimization role. For instance, if the goal is to find out from where people are visiting a website, not storing the referrer with other visitor data, but rather keeping a running tally as they visit, would preserve the functionality while increasing visitor privacy.

ii. Obfuscation

Obfuscation refers to the hiding of data or obscuring its meaning. There are several techniques for doing this. Among the most common is adding noise to the data or randomizing the data set. Randomization can help reduce linkability that can be obtained based on ordering. As an example of the linkability of non-randomized data sets, Shannon Richardson was caught by the U.S. Postal Service mailing a ricin-laced letter because they were able to determine (based on the order in which mail was collected) what time her letter entered the mail stream. This was important because she blamed her husband, who was, fortunately for him, at work at the time.[40] Differential privacy provides a formalized method of adding noise that still preserves the meaningful statistics about the data set.[41]

Hashing and encryption are other forms of obfuscation which hide the original meaning of the information. The latter is accessible to those with secret knowledge and the former still useful for those needing to prove they know the original meaning or to ensure consistency to preserve some degree of data integrity.

iii. Security

As previously noted, security engineering is a rather mature field with significant research supporting the craft. Security can be supportive of policies and used to restrict the access and use of information. Point controls include identity and access management systems, encryption, authenticating use, auditing and logging and data loss prevention systems which monitor the movement of data outside specified boundaries. Threats are constantly changing and evolving and these threats must be analyzed as to the sufficiency of existing security controls and how privacy may be affected.

[40] Ron Nixon, "US Postal Service Logging all mail for law enforcement" NY Times July 3, 2010 Available at http://www.nytimes.com/2013/07/04/us/monitoring-of-snail-mail.html?_r=0

[41] See Microsoft Corp. (2012), "Differential Privacy for Everyone" and Cynthia Dwork, "The Promise of Differential Privacy. A Tutorial on Algorithmic Techniques." (2011, 52nd Annual IEEE Symposium on Foundations of Computer Science) at http://research.microsoft.com/apps/pubs/default.aspx?id=155617

VIII. Conclusion

This paper surveyed the emerging discipline of privacy engineering. Privacy engineers require multidisciplinary knowledge and skills. To be effective, they need to have an understanding of both technical and non-technical considerations. Privacy engineers are tasked with managing *risks*, and there are various risk models that they can adopt, some emphasizing Fair Information Practice Principles and legal compliance, others focusing on harms and contextual integrity. Privacy engineers must then apply systematic risk analyses, using tools such as privacy impact assessments, to measure and quantify identified risks. Finally, privacy engineers must design controls to mitigate those risks, including privacy-respecting architectures, effective privacy policies, and a range of data management methods including minimization, anonymization, aggregation, and the use privacy-enhancing technologies.

With the shift from industrial manufacturing to knowledge creation and service delivery, the value of information and the need to manage it responsibly have grown dramatically. At the same time, rapid innovation, global competition, and increasing system complexity present profound challenges for informational privacy and data protection.

While we would like to enjoy the benefits of innovation – new conveniences and efficiencies – we must also preserve freedom of choice and personal control over personal data flows. Always a social norm, privacy and data protections have nonetheless evolved over the years, beyond being viewed solely as a legal compliance requirement, to being recognized as a market imperative and critical enabler of trust and freedoms in our present-day information society.

There is a growing understanding that innovation and competitiveness must be approached from a “design-thinking” perspective – namely, a way of viewing the world and overcoming constraints that is at once holistic, interdisciplinary, integrative, creative, innovative, and inspiring.

Privacy, too, must be approached from the same design-thinking perspective. Privacy and data protection should be incorporated into networked data systems and technologies **by default**, and become integral to organizational priorities, project objectives, design processes, and planning operations. Ideally, privacy and data protection should be embedded into every standard, protocol, and data practice that touches our lives, by design. This will require skilled privacy engineers and common methodologies and tools, but will be well worth the effort. The future of privacy, and in turn freedom, may well depend on it.

www.privacybydesign.ca

Office of the Information and Privacy Commissioner,
Ontario, Canada
2 Bloor Street East
Suite 1400
Toronto, Ontario
Canada M4W 1A8
Telephone: 416-326-3333
Fax: 416-325-9195
E-mail: info@ipc.on.ca
Website: www.ipc.on.ca

The MITRE Corporation
@MITREcorp
Website: http://www.mitre.org

Enterprivacy Consulting Group
@privacymaverick.com
Website: http://blog.privacymaverick.com

The information contained herein is subject to change without notice. MITRE Corporation, Enterprivacy Consulting Group and the IPC shall not be liable for technical or editorial errors or omissions contained herein.

Web site: www.ipc.on.ca
Privacy by Design: www.privacybydesign.ca

January 2014