Developing Foundations for Accountability Systems: Informational Norms and Context-Sensitive Judgments

Robert H. Sloan
University of Illinois at Chicago
Dept. of Computer Science (MC 152)
Chicago, IL 60607-7053
sloan@uic.edu

Richard Warner
Chicago-Kent College of Law
565 W. Adams St
Chicago, IL 60661-3652
rwarner@kentlaw.edu

ABSTRACT

Adequately protecting informational privacy in an increasingly interconnected world poses two problems. What are the appropriate privacy polices? And, how should one ensure compliance with them?

Accountability systems are an attractive solution to *both* problems. Current work on accountability systems assumes a generally accepted set of privacy rules for the subsequent use of information, and has focused on developing a formal representation of a process for the use of information. Our focus is on fundamental policy issues that arise in developing the models of the privacy rules themselves. This focus leads to the suggestion that accountability systems can be used, not only to enforce compliance with a given set of rules but also to resolve conflicts among conflicting sets of rules. So far, accountability systems have modeled unrealistically simple privacy rules. While this may be an appropriate first step toward more complex systems, we need to define the realistic target at which accountability systems should ultimately aim if adequate systems are eventually to be developed. We specify a number of hurdles to developing accountability systems that adequately constrain the use of information. Some of the problems are wholly non-technical; some are of a mixed nature, part social science or public policy and part technical. The unifying theme is the role of informational norms in ensuring adequate informational privacy.

Categories and Subject Descriptors

K.4.1 [**Computers and Society]** Public Policy Issues – *privacy, regulation.*

General Terms

Management, Security, Legal Aspects.

Keywords

Accountability, norms, privacy, information accountability, accountability systems.

GTIP 2010, Dec. 7, 2010, Austin, Texas, USA.

1. INTRODUCTION

How does one adequately protect informational privacy in a world that is increasingly interconnected but still fragmented by differing laws, customs, and world views? The question divides into two. What are the appropriate privacy polices? And, how should one ensure compliance with them? In a widely cited 2008 article in *Communications of the ACM,* Weitzner et al. [15,16] offered an attractive solution to the second question. They assume a generally accepted set of privacy rules and propose a tracking process for the use of information that would create an incentive to abide by the rules by making uses transparent. In short, instead of (or in addition to) access control, Weitzer et al. propose giving everybody the ability to determine, after the fact, who accessed which information. Call any such system an accountability system. We suggest (in Section 3.1.3) that, despite the problematic nature of the initial assumption of an accepted set of rules, the development of accountability systems can be an important step toward answering the question of what privacy polices ought to be adopted. In considering such systems, we focus exclusively on commercial interactions; they raise complex and important issues that have not been as extensively examined as governmental intrusions into privacy.

We contend that Weitzner et al.'s accountability system faces serious difficulties; our point, however, is not to reject their system but to develop it. Given the vast and ever-increasing amount of information available over the Internet, there seems little alternative to some form of automated checking for compliance with privacy requirements. We hope that accountability systems can provide the necessary automated assessment. They are unlikely to do so however, without an adequate foundation in both formal models and public policy issues, and, as Jagadeesan et al. note, "the accountability approach to security lacks general foundations for models and programming" [4]. Jagadeesan et al. develop formal foundations in two steps. They first describe an operational model in which privacy policies define what information may and may not be shared among various agents; the model is based on Communicating Sequential Processes (CSP) [2], and the traces of the various agents' processes. They then provide algorithms that an auditor can use to check a certain form of compliance with rules. Like Weitzner et al., Jagadeesan et al. simply (and rightly given their purposes) assume that appropriate rules exist.

Here we also contribute to the development of foundations for accountability systems. Weitzner et al. give a broad outline of research problems that must be solved in order to develop accountability systems, concentrating on technical problems;

Jagadeesan et al. focus in on some important formal-logic verification problems arising from the development of accountability systems. Our contribution is not, however, to the formal or technical foundations, but to the equally fundamental public-policy issues that arise in developing the models of the rules that one then formally represents. To this end, we combine work in computer science on accountability systems with the work of social theorists (e.g., Helen Nissenbaum) on the critical role of norms in ensuring adequate informational privacy [9]. Our key claim is that the privacy rules relevant to developing accountability systems are, for the most part, informational norms. Informational norms are social norms that constrain the collection, use, and distribution of personal information.

We sketch a number of problems that must also be solved to develop accountability systems. Some of the problems are wholly non-technical; some are of a mixed nature, part social science or public policy and part technical. The unifying theme is the role of informational norms in ensuring adequate informational privacy. The problems are:

- Developing machine-readable forms of subtle, nuanced privacy rules.
- Ensuring the optimality of trade-offs made by privacy rules.
- Developing contextually sensitive reasoning tools.
- Developing new norms where relevant norms currently do not exist.
- Creating incentives for businesses and individuals to use accountability systems.
- Resolving inconsistencies in norms from population to population.

2. INFORMATIONAL NORMS AND INFORMATIONAL PRIVACY

Why think that the rules and polices relevant to developing accountability systems are, for the most part, informational norms? Our answer in summary form: (1) the relevant rules should implement generally accepted trade-offs between informational privacy and competing concerns; (2) the rules that do so are for the most part informational norms. We begin the argument for (1) by clarifying the notion of informational privacy. Informational privacy is a matter of control. It is "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [18]. Privacy advocates insist—rightly—that a significant degree of control over personal information is essential to "protecting intimacy, friendship, individuality, human relationships, autonomy, freedom, self-development, creativity, independence, imagination, counterculture, eccentricity, creativity, thought, democracy, reputation, and psychological well-being." [13] Anyone concerned with such ends has a strong incentive to avoid activities that significantly reduce informational privacy. A lack of constraints on the use of initially voluntarily disclosed information seriously threatens to reduce informational privacy and hence creates a strong incentive to withhold information. Thus, constraints are called for if the Internet is to reach its full information-sharing potential. Defining the constraints is no simple task, however. The broad use of information yields significant benefits, including increased availability of relevant information, increased economic efficiency, and improved security [5]. Therefore any acceptable set of privacy rules must balance the benefits against the loss of information privacy.

An accountability system must incorporate such rules. If the rules fail to adequately balance informational privacy against competing concerns, then the accountability system will encode rules that yield unacceptable results. Moreover, the rules must be generally accepted rules. If not, the accountability system is not a representation of people's preferences in regard to privacy but an attempt to impose a view about what ought to be private. We assume that the goal of accountability systems is to represent privacy preferences, not to legislate them. There are three plausible candidates for generally accepted rules that adequately balance competing concerns: legal rules; the rule that the information may only be used in ways to which the subject of the information has consented; and, informational norms. We will discuss each in order, and argue that the last dominates the field.

2.1 Legal Rules

There are not currently laws or regulations (at least in the United States) that would allow accountability systems to adequately constrain the use of information [12]. Current laws place relatively few restrictions on private sector processing of personal information; moreover, proposals for further regulation encounter considerable controversy over precisely how to balance privacy against competing concerns. As the privacy advocate James Rule notes, "[w]e cannot hope to answer [complex balancing questions] until we have a way of ascribing weights to the things being balanced. And, that is exactly where the parties to privacy debates are most dramatically at odds." [11]

We conclude that legal regulation (at least in the United States) does not offer, and is not likely in the future to offer, a sufficiently comprehensive array of rules to allow accountability systems to adequately constrain the use of private information.

2.2 Consent Requirements

Consent requirements come in two forms. The first is the requirement that businesses present consumers with relevant information in an understandable fashion and then secure (in some specified fashion) agreement to proceeding with the transaction. The second are Platform for Privacy Preferences (P3P)-like approaches that provide a way for each Web user to give or withhold consent to requests to collect information about them. We consider P3P-like approaches first. Weitzner et al. point out a crucial flaw:

> A fully-implemented P3P environment could give Web users the ability to make privacy choices about every single request to collection information about them. However, the number, frequency and specificity of those choices would be overwhelming especially if the choices must consider all possible future uses by the data collector and third parties. Individuals should not have to agree in advance to complex policies with unpredictable outcomes. [16]

The unpredictability problem is actually worse than the above passage suggests. Weitzner et al. confine their attention to information that explicitly identifies one as the individual whom

the information describes; they do not consider anonymized information; however, given the power of reidentification algorithms, one must be able to predict future uses even of anonymized information [6-8].

Even if we put aside the "overwhelming choice" problem, the proposal is still problematic. Assume consumers could obtain and understand all the relevant information; it would still be unlikely that the overall pattern of consent would determine a socially optimal trade-off between privacy and competing concerns. Consider an analogy. At least before the era of the Web, comprehensive telephone books usefully facilitated communication. Suppose, however, that while most people preferred telephone books with most other people's numbers in them, a majority also preferred not to have their individual numbers listed. In such a case, if consent were required to list a number, reasonably comprehensive telephone books would not have existed. A similar suboptimal outcome might well result from a workable implementation of P3P. People may withhold too much information. "There is often little individual incentive to participate in the aggregation of information about people, [yet] an important collective good results from the default participation of most people." [14]

The same objections apply to requiring businesses to present relevant information and then to secure agreement to proceeding with the transaction. Consumers have to assess complex policies with unpredictable consequences [18], and a socially optimal trade-off between privacy and competing concerns is unlikely.

2.3 Informational Norms

Neither legal rules nor consent requirements are likely to yield rules that adequately constrain the subsequent use of previously disclosed information. Informational norms can—and do—play this role. As Nissenbaum notes, informational norms

> [g]enerally . . . circumscribe the type or nature of information about various individuals that, within a given context, is allowable, expected, or even demanded to be revealed. In medical contexts, it is appropriate to share details of our physical condition or, more specifically, the patient shares information about his or her physical condition with the physician but not vice versa; among friends we may pour over romantic entanglements (our own and those of others); to the bank or our creditors, we reveal financial information; with our professors, we discuss our own grades; at work, it is appropriate to discuss work-related goals and the details and quality of performance. [9]

Informational norms are instances of the following pattern: a person or entity may collect, use, and distribute information only as is appropriate for the social role the person or entity is playing. "Appropriateness" is determined contextually. Over a wide range of cases, group members share a complex of values that leads them to more or less agree in their particular contextual judgments of appropriateness. Understanding privacy via norms yields a far more context-sensitive approach than merely thinking of private information as personally identifiable information; privacy norms, for example, allow pharmacists to obtain personally identifiable information about the drugs you are taking, but not about whether you are happy in your marriage. The approach also yields a much broader concept of privacy than the typical industry understanding of private information as information protected by legislation and compliance requirements. Three further points are in order. Each introduces an assumption that we will make in our further discussion in Section 3.

First, Weitzner et al. assume that the rules governing the subsequent use of voluntarily disclosed information are encodable in a machine-readable form. However, the relevant rules are (for the most part at least) informational norms whose application is determined by value-laden, contextually varying judgments of appropriateness, and it is unclear whether such norms can be encoded in a machine-readable form. Current formalizations of informational norms simply sidestep this problem, as Barth et al. [1] illustrate. They use linear temporal logic to provide a formal representation of norms. They illustrate their approach with an example drawn from the 1999 Gramm–Leach–Bliley Act, which sets privacy rules financial institutions must meet when processing customer information. However, as Barth et al. note in examining the Gramm–Leach–Bliley Act, some rules concern "affiliates" of financial institutions and "non-public personal information." There is a "complex definition of which companies are affiliates and what precisely constitutes non-public personal information." Determining whether a company is an affiliate requires judging whether it "controls, is controlled by, or is under common control with another company," and determining whether information is non-public personal information requires applying the following definition. Non-public personal information is "personally identifiable financial information (i) provided by a consumer to a financial institution; (ii) resulting from any transaction with the consumer or any service performed for the consumer; or (iii) otherwise obtained by the financial institution," and this definition is further qualified by complex exceptions specified in the Act. As Barth et al. note, "*Our formalization of these norms sidesteps these issues* [emphasis added] by taking the role affiliate and the attribute npi [non-public personal information] to be defined exogenously: the judgments as to which companies are affiliates and which communications contain npi are made in the preparation of a trace history [of the relevant communications to be formalized]" [1]. A similar point holds for various access control frameworks and various "privacy languages" that have been proposed, including: RBAC, XACML, EPAL, and P3P. The formal structures do not provide any means to capture the fact that the application of a term may vary as contextual judgments vary. In the rest of this article we assume that the relevant information norms can be encoded in an adequate formal, machine-readable manner. We suggest that the fact that the application of terms vary with varying contextual judgments may in some cases be better addressed in the development of a program that reasons about how to apply norms in a particular context instead of in the representation of norms themselves.

Second, as we emphasized earlier, norms define trade-offs between informational privacy and competing goals; however, the trade-offs may be poor ones. We will not address this problem; instead, we assume norm optimality: all informational norm trade-offs are at least as well justified as any alternative.

Third, in many cases, the rapid advance of computing has outstripped the relatively slow evolution of social norms; hence, information processing is often unconstrained by relevant norms.

Cloud-computing services provide one example. The services maintain data generated by users' activity on the services' servers. No norm defines what a cloud-computing service provider may do with the information it processes. The service providers vary significantly in the extent to which their information processing invades informational privacy; moreover, as the sharp controversy over cloud-computing privacy shows, there is no agreement on a norm [3]. As with machine readability, and norm optimality, we put problems about the existence of norms aside. We assume norm completeness: all information processing is governed by generally accepted informational norms.

3. THE TRACE-BACK PROCESS

Weitzner et al.'s accountability system is premised on three claims: (1) there are privacy rules governing the use of information; (2) those rules adequately balance informational privacy against competing concerns; and (3) after-the-fact accountability ensures an adequate incentive to abide by those rules. The assumptions of machine readability, norm optimality, and norm completeness guarantee the fulfillment of (1) and (2). However, even given these strong assumptions, (3) is problematic. It is implicit in the notion of accountability that there must be a *trace-back process*, that is, a process by which an auditor, and perhaps any end user, can verify that all the uses of some piece of information were in compliance with the policy rules. Weitzner et al. propose four parts for this process. We consider each in turn.

3.1.1 Part One: policy reasoning tools.

Weitzner et al. observe that

> Accountable systems must assist users in seeking answers to questions such as: Is this piece of data allowed to be used for a given purpose? Can a string of inferences be used in a given context, in light of the provenance of the data and the applicable rules? It seems likely that special purpose reasoners, based on specializations of general logic frameworks, will be needed to provide a scalable and open policy reasoner. [16]

They note that "an initial application of these reasoners has been implementation of policy aware access control that enable standard web servers to enforce ruled-based access control policies specifying constraints and permissions with respect to the semantics of the information in the controlled resources and elsewhere on the Web." [16] The reasoning involved in such systems is not, however, remotely like the reasoning about informational norms. Here is a typical example of reasoning about rule-based access [17].

Alan: (1) If X is AC rep of Y, X can delegate W3C membership rights in Y. (2) Kari is AC rep of Elissa.

Kari: (1) If X is employee of Elissa, X has W3C membership rights. (2) Tina is employee of Elissa.

Tina: I have W3C membership rights. Proof: Alan1, Alan2, Kari1, Kari2.

Compare the reasoning required to apply informational norms. Consider the norm that a wine retailer may process information only in ways appropriate to a wine retailer. Suppose the wine store collects and analyzes information to determine the sexual orientation of its customers. One must reason from this fact and the norm to the conclusion that the information processing is or is not permissible under the norm. This requires determining if the processing is "appropriate." Judgments of appropriateness are a function of applying a complex of shared values and attitudes in a particular context. As the example of non-public information discussed in Section 2.3 illustrates, such judgments involve a degree of complexity and context-sensitivity far beyond the relatively simple judgments about access illustrated by the Alan-Kari-Tina example. We are still a long way from developing a reasoning system that can, for example, reliably match the judgments of a trained lawyer about whether a particular piece of data is public or non-public information under Gramm-Leach-Bliley.

3.1.2 Part Two: policy aware transaction logs.

At "endpoints" a log will be created of "information usage events" which are "relevant to current or future assessment of accountability to some set of policies." [16]

We note in passing that it is unclear about what an endpoint is—an individual computer (or network), an ISP? Different choices mean different allocations of the burden of storage and security (including ensuring legitimate access to the information). Our main concern is with the notion of a "usage event." There are obvious problems if it means logging every transaction everyone makes everywhere. Anything less, however, would seem to give Weitzner et al. less than they desire. They ask one to consider the following scenario:

> Alice is the mother of a three-year old child with a severe chronic illness that requires long-term expensive treatment. She learns all she can about it, buying books online, searching on the Web, and participating on online support parent-support chat rooms. She then applies for job and is rejected, suspecting it's because a background check identified her Web activities and flagged her as high risk for expensive family health costs. [16]

They assume that "the decision to deny Alice the job . . . [was an] inappropriate use of that information" [16]. We do not see how such scenarios can be prevented unless on logs every transaction everyone makes everywhere. Many decision makers in hiring are the sorts of people who do some of their work from home, and would naturally do a web search just to see what they might learn about a finalist for a job. Thus we would need logs not only for the computers at the company that was considering hiring Alice, but also for the computers in the homes of the company's employees who make hiring decisions.

3.1.3 Part Three: the policy language framework.

They acknowledge that global compatibility in the language used to create logs is unlikely, and indeed that the rules—the informational norms—will vary from group to group. They envision a resolution mechanism like the judicial mechanisms for resolving jurisdictional questions and conflicts of law. But the analogy is more apposite than they realize. The judicial mechanism is slow, expensive, and fraught with controversy arising from the pressures of globalization and the Internet; moreover, it requires highly trained, human decision makers.

Accountability systems may nonetheless make an important contribution to the resolution of conflicts. The explicit, machine-

readable formulation of norms and the development of machine-representable, context-sensitive reasoning facilitate the detailed identification of similarities and conflicts. This can provide the input into "second-order" accountability systems designed to resolve conflicts as they arise. Such accountability systems promise a solution to adequately protecting informational privacy in an increasingly interconnected but still fragmented world.

3.1.4 Part Four: accountability appliances.

Weitzner et al. envision a collection

> of accountable appliances throughout the system that communicate through Web-based protocols. Accountability appliances would serve as proxies to data sources, mediating access to the data, and maintain provenance information and logs of data transfers. They could also present accountability reasoning in human-readable ways, and allow annotation, editing, and publishing of the data and reasoning being presented. [16]

While they do not say so explicitly, we assume that they envision a collection of private, non-governmental accountability appliances. The critical question is, again, how to incentivize or compel businesses to use the appliances. There is reason to doubt that businesses will do so voluntarily. Collecting personal information about customers confers a significant competitive advantage on the business; consequently, the more aggressively a business's competitors harvest customer information, the more of an incentive the business has to do so as well. As Privacy International notes in a 2007 report,

> In contrast to the 1990's vision of the Internet, in which strong privacy could become a market differentiator, the reality in 2007 is that all major Internet players may move to establish a level of user surveillance that results in little or no choice for Internet users and relatively few meaningful privacy mechanisms. Market domination by a handful of key players will ensure that without care, a race to the bottom will evolve during the immediate future. [10]

4. Conclusion

We by no means deny that accountability systems have a role to play in ensuring adequate informational privacy. Given the vast and ever-increasing amount of information available over the Internet, there seems little alternative to some form of automated checking for compliance with privacy requirements. For accountability systems to play this role, several problems must be overcome. Weitzner et al. laid out many of the technical problems, such as architectural and scalability issues. Here we have presented a number of additional problems based on public-policy considerations.

First, machine-readability: An adequate machine-readable representation of informational norms must be developed. This will require addressing the fact that the applications of crucial terms vary as the context varies, unless the entire issue is dealt with in the development of reasoning tools.

Second, contextually-sensitive reasoning tools: Human reasoning about the application of informational norms involves context-sensitive judgments. A context-sensitive reasoning program that can make similar judgments is required.

Third, lack of norms: As a result of rapid advances in information processing technology, there are no appropriate informational norms that constrain businesses' information processing across a wide range of cases. The lack of informational norms blocks the use of accountability centers precisely where they are most needed—where rapid advances in information processing technology have facilitated both novel forms commercial and social interaction and the collection, analysis, and distribution of vast amounts of information concerning those involved in such interactions. It would be a striking achievement of great importance if accountability systems not only constrained the use of information in light of existing norms, but also contributed to the generation of new norms by revealing patterns of interaction between consumers and businesses. Perhaps Weitzner et al.'s "accountability appliances" could play a role here. As they note, "accountability appliances could . . . present accountability reasoning in human readable ways, and allow annotation, editing, and publishing of the data and reasoning presented" [16]. It is worth investigating whether such interaction could contribute to the development of norms.

Fourth, lack of incentive: We doubt that private businesses have an adequate incentive to use privately maintained accountability centers. Accountability centers may have to be developed under the assumption that appropriate legal regulation will mandate their use.

Fifth, resolution of inconsistencies: Not only will the language in which norms are encoded vary from region to region, so will the norms themselves. We can address this problem through second-order accountability systems designed to resolve conflicts as they arise. The diversity of cultures, traditions, and conceptions of privacy suggests that conflict sets of rules, rather than agreement on a single set of rules, is a permanent condition. If so, second-order resolution of conflicts as they arise is the solution. A related problem is sub-optimal norms, norms that are not as well justified as any alternative. The second-order examination and resolution of conflict may suggest improvements in particular first-order norms.

Sixth, data storage: If accountability systems are to adequately constrain the use of information, their trace-back systems evidently require the storage of an immense amount of information. It is unclear how this is to be accomplished.

5. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-0959116.

6. REFERENCES

[1] Barth, A. et al. 2006. Privacy and contextual integrity: Framework and applications. *Proceedings of the IEEE Symposium on Security and Privacy* (2006), 184–198.

[2] Brookes, S.D. et al. 1984. A theory of communicating sequential processes. *Journal of the ACM (JACM)*. 31, 3 (1984), 560–599.

[3] Gellman, R. 2009. *Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing*. World Privacy Forum.

[4] Jagadeesan, R. et al. 2009. Towards a Theory of Accountability and Audit. *ESORCS'09, 14th European Symposium on Research in Computer Security* (2009), 152–167.

[5] Kang, J. 1998. Information privacy in cyberspace transactions. *Stanford Law Review*. 50, 4 (1998), 1193–1294.

[6] Narayanan, A. and Shmatikov, V. 2008. Robust De-anonymization of large sparse datasets. *Proceedings of the IEEE Symposium on Security and Privacy* (2008), 111–125.

[7] Narayanan, A. and Shmatikov, V. 2009. De-anonymizing social networks. *Proceedings of the IEEE Symposium on Security and Privacy* (2009), 173–187.

[8] Narayanan, A. and Shmatikov, V. 2010. Myths and fallacies of personally identifiable information. *Commun. ACM*. 53, 6 (2010), 24–26.

[9] Nissenbaum, H. 2004. Privacy as contextual integrity. *Washington Law Review*. 79, 1 (2004), 119–158.

[10] Privacy International 2007. *A Race to the Bottom: Privacy Ranking of Internet Service Companies*. Privacy International.

[11] Rule, J.B. 2007. *Privacy in Peril: How We are Sacrificing a Fundamental Right in Exchange for Security and Convenience*. Oxford University Press.

[12] Schwartz, P.M. 2000. Internet Privacy and the State. *Connecticut Law Review*. 32, (2000), 815–859.

[13] Solove, D.J. 2008. *Understanding Privacy*. Harvard University Press.

[14] Walker, K. 2001. Costs of Privacy, The. *Harvard Journal of Law & Public Policy*. 25, (2001), 87–128.

[15] Weitzner, D.J. et al. 2007. *Information Accountability*. Technical Report #2007-034. MIT Computer Science and Artificial Intelligence Laboratory.

[16] Weitzner, D.J. et al. 2008. Information accountability. *Commun. ACM*. 51, 6 (2008), 82–87.

[17] Weitzner, D.J. et al. 2006. Creating a Policy-Aware Web: Discretionary, Rule-Based Access. *Web and Information Security*. IRM Press. 1–31.

[18] Westin, A. 1967. *Privacy and Freedom*. Atheneum Press.