Article

new media & society
0(0) 1–17

DOI: 10.1177/1461444811426741
nms.sagepub.com
SAGE

Privacy by design: Networked computing, 1969–1979

Sandra Braman
University of Wisconsin-Milwaukee, USA

Abstract
Discourse analysis of the technical document series that records the internet design history, the RFCs, shows that those involved during the first decade saw privacy as a multi-dimensional and interactive problem requiring use of a suite of solutions at the network, individual, and data levels that had to take into account the need to balance privacy against experimentation and innovation. Internet designers were sophisticated in their pragmatic thinking about privacy when evaluated vis-a-vis theoretical developments since that time, viewing privacy as a contextual matter involving boundary setting, and using information architecture and metadata as tools for privacy protection. Those in the social science and legal communities think about the privacy effects of communication on humans, while those in the technical design community must focus on privacy as a set of logistical problems. Bringing these diverse communities into a single conversation can considerably enrich and strengthen the work of all.

Keywords
ARPANet, information architecture, information policy, innovation, internet, network architecture, network design, privacy, RFCs, sociotechnical

If you have a secret, don't keep it on the ARPAnet

(Brian Harvey, 1975, RFC 686: 3)

Technical designers of the internet quickly realized what is now common knowledge: it is hard to protect privacy online. The effort to protect privacy was a constant from the issuance of the first US government contract to link computers at different sites in 1969. About 17 percent of the 718 documents published through the close of 1979 in the technical

Corresponding author:
Sandra Braman, Department of Communication, University of Wisconsin-Milwaukee, PO Box 413, Milwaukee, WI 53201, USA
Email: braman@uwm.edu

document series that records the history of the design process – the Requests for Comments, or RFCs – deal with privacy,[1] the most frequently discussed social policy issue. Developments during the first decade of work – the 'framing years' (Braman, 2011) – were particularly influential because the initial decisions for what began as ARPANet[2] created the conditions under which 21st-century online threats to privacy became possible (Blumenthal and Clark, 2001; Denardis, 2009).

This article looks at the treatment of privacy issues in the RFCs from 1969 to 1979 using a discourse analysis that involved a *comprehensive* and *inductive* reading of the documents. Both features of this method were critical. Because about one third of the items identified through word search had only a spurious relationship to the subject as a policy issue and privacy issues were often discussed without using obvious terms, it was essential to read every line of every document. Because policy analysis of technical documents is a secondary reading requiring significant sociotechnical boundary work, the analysis had to be inductive and, often, conceptual, in order to elicit the privacy implications from descriptions of technical problems and their possible resolutions.

The RFCs are worth studying as a discourse because they are considered by insiders to be the 'documents of record' for the technical history of the internet (Leiner et al., 2010). The series was launched by a few graduate students not long after they began working together to link the computers of their geographically distant sites. The goal was to share ideas and information within the quickly growing community. At launch, the series was insistently informal and welcomed all comers; over time, the publication process became formalized. Governance institutions such as the Internet Engineering Task Force (IETF) and the Internet Corporation for Assigned Names and Numbers (ICANN) developed through the RFCs. Today, the texts are freely available online, hosted by the IETF at www.ietf.org.

This research is part of a larger project analyzing the treatment of legal and policy matters in the RFCs through 2009 (Braman, 2010).[3] To fully grasp the origins of today's online privacy issues, however, a detailed understanding of the early thinking that shaped the network is necessary. Technical thinking of the era was so different from that of social scientists and legal scholars that these findings also include ideas about how to design privacy protection policies for technical environments, and specific protection techniques, that can valuably be added to the toolkit of today's policy makers. Legal approaches to protecting privacy fail when they are incomplete, and when they do not take into account the actual nature of the technologies involved; social scientists will see early foreshadowings of some ideas with great currency today, and encounter others that may stimulate new research and social theory.

Empirical research on and theorizing about privacy have become ever-more important scholarly enterprises. Because of space limitations, unfortunately there is room here only for a few outstanding exemplars from the literature. Privacy is inherently political (Branscomb, 1986; Star and Ruhleder, 1996), involving the very boundary-defining activities (Petronio, 2002) that are so flexible and complex in the networking environment. The same user can have competing privacy interests (Case, 2000); determining which dominates in any given circumstance is a contextual exercise (Nissenbaum, 2004). Privacy suffers from 'policy drift' because it lacks an organized constituency, and because practice tends to dominate over explicit decision making (Smith, 1994). The economic costs of privacy invasions are difficult to quantify, though not long after the period reported upon here

privacy protection became a marketable service (Auerbach, 1983). Technological innovations introduce new vulnerabilities that are often difficult to foresee because they are unfamiliar or result from such complex interactions that they may be unknowable until they occur.

The internet design community addressed privacy issues at the network, the individual, and the data levels during 1969–1979. They understood that privacy is interdependent; effective protection requires successful use of the entire bundle of privacy protections across those three levels. They understood that privacy protections themselves can require privacy. Privacy conditions change, requiring renewed attention with every innovation or change in the system. Other ideas from the period can be, or already have been, brought into recent use by practitioners, if not by policy-makers or theorists.

The article begins by looking at privacy as a network design problem. It goes on to look at arguments for, and then against, privacy protection at each level before examining the techniques for protecting privacy discussed in the RFCs.

Privacy as a design problem

The internet is not the first communication technology to raise privacy concerns. During the 19th and early 20th centuries, what were experienced as invasions of privacy by reporters for print newspapers so angered people that violence against the press resulted (Nerone, 1994). The telephone, introduced in the 1880s, allowed individuals into the home who would not have been permitted to enter previously (Marvin, 1988). Between 1930 and 1950, police take-up of the combination of cars and radios diminished citizen privacy (Dandeker, 1990). Protecting the privacy of networked communications was a policy issue beginning with the telegraph; it was one of the first topics addressed in 1865 by the organization that became the International Telecommunications Union (ITU) (Codding, 1972). Studies of computing in the 1950s and 1960s concluded that new technologies exacerbated privacy as a social problem (Kling, 1980), and the issue was covered in a series of articles on the information society in the most prestigious economic newspaper in Japan in 1969 (Ito, 1991).

The ubiquity and complexity of the internet make privacy a networked communication policy issue of central importance. During the period discussed here, governments around the world were thinking deeply about privacy issues and reviewing their own laws and practices. The internet is a 'network of networks' (RFC 1122) that was international in intention from the start and in reality by the mid-1970s (Braman, in press). It was US law, though, that provided the legal context for ARPANet/internet designers of the 1970s; all but 5 of the 718 documents in the series published by the close of 1979 were authored by individuals employed by organizations headquartered in the USA.

Privacy concerns generated by new types of databases for census information and labor statistics led to a series of studies funded by private foundations as well as the government (US Department of Health, Education, and Welfare, 1973; Privacy Protection Study Commission, 1977; Westin and Baker, 1972). All of these warned that privacy problems would become more serious when databases became networked. Books by Alan Westin (e.g. 1970) and by Arthur Miller (1971) popularized the issue. Other pertinent developments during the period included passage of the Privacy Act in 1974 and the introduction of principles for fair digital information practices (Regan, 2008; Trubow, 1989).

Active contributors to the internet design process, such as the RAND Corporation that works so closely with the US government, had been working on digital privacy problems before becoming involved in building a working network (RFC 243). The difference between reaching a consensus on privacy protection as a general principle and achieving agreement on actual techniques to be used immediately became evident, as did the need to balance privacy against maximizing the capacity for experimentation and innovation (RFC 195). It was also recognized, however, that techniques developed to protect privacy could serve additional technical and social functions (RFC 269).

Privacy first appeared in the RFCs in a description of variations in practice (RFC 109). By 1972, the need for log-in privacy was so widely accepted that the use of passwords showed up without comment in an example of a protocol (RFC 307), though the same could not yet be said for masking such information (RFC 318). As a categorization system for protocols developed, privacy (RFC 750), and then security (RFC 753), were identified as running topics. The issue inevitably arose in discussions about access (RFC 487). Every site was asked to provide information pertinent to how it protected privacy at the points of log-in, protection of online activity, storage, and output in a survey intended to provide support to remote users (RFC 364). In 1978, the ability of sites to send mail to unknown users (RFC 751) – which can be experienced as invasions of privacy – was tested.

Throughout the decade, privacy-related concepts were further articulated (e.g. RFC 435), and additional vulnerabilities were described (e.g. RFC 666). Still, some felt that privacy was not receiving enough attention, constantly being postponed to be dealt with 'later' – even though there was evidence that privacy problems were far worse than was being generally acknowledged (RFC 602). Various privacy issues were conflated in a way many believed unuseful (RFC 501). A distinction between security and privacy was acknowledged but never clarified conceptually; for technical decision-makers, the logistical problems were the same. Arguments both for and against protecting privacy were in play.

Arguments for protecting privacy

Arguments for including privacy protections in internet design were presented from the perspectives of the network as a whole, of individual hosts, and of the user. Distinctions among the three levels were clear to network designers (see e.g. RFC 610).

Privacy at the network level

Four arguments for protecting privacy at the network level emerged during the first decade of the RFCs. There was appreciation of the critical role of privacy as essential to network integrity, an affordance for resource sharing through the network, a support for accounting systems, and an element of professionalization.

Protection of network integrity. The need to protect network integrity was so important that one strongly supported proposed protocol (RJE) was abandoned because of its weaknesses on this front (RFC 725). The general need to ensure network integrity (RFC 62) became unbundled into a number of distinct problems as the design effort progressed. The importance of trust, today widely recognized as key to success in any type of networked activity,

was discussed by network designers as early as 1971 (RFC 98). Users of protected file systems, they pointed out, must be able to have confidence that servers can correctly identify remote users (RFC 114). Authors writing from a national security perspective employed the trickle-down argument: developing privacy protections to military specifications would result in enhanced privacy for all (RFC 316).

As has been the case throughout the internet design process, humans and 'daemon' (computing process) users were treated separately (Braman, 2011). Although policy-makers and social scientists are not accustomed to applying the concept of privacy to non-humans, network designers felt that, at least under certain conditions, socket names deserved privacy (RFC 54) and that server processes needed to be able to securely exchange those names in order to establish a trusted connection (RFC 430). It was unclear which identifier for a computing process should be used when determining access rights (RFC 501), or how a server should determine whether or not any given process required a distinct log-in process for identity verification (RFC 555).

Enabling resource sharing. Two types of resource sharing were discussed during the first decade of the internet design process – computational resources and data. We tend to think of sharing as a human activity, but for internet designers of the 1970s resource sharing was a form of 'interprocess' communication linking specific resources to particular processes (RFC 61). However, it was also understood that privacy for human users was essential to what was, at the time, referred to as 'indirect' use of networked computers (RFC 114) – computing involving two or more machines and/or undertaken at a distance.

Privacy was one of six broad areas identified as crucial for data sharing in 1971 (RFC 146), equal in importance to manipulating files across systems, logically restructuring data in response to queries, standardizing data management across computers, and keeping duplicate copies of a database consistent. As reliance upon networked databases grew, so did the sense of urgency regarding the need to protect privacy (RFC 340), even though some users continued to prefer private connections for batch processing (RFC 647).

Accounting. As soon as ARPA-funded host sites opened themselves up to users without US government support, the question of who was doing what became a pressing accounting problem. Authentication of user identities was necessary for billing purposes, if for nothing else (RFC 136); without it, the cost of providing services would become system overhead at a level unsustainable for serving hosts (RFC 487). Passwords were an obvious means of requiring user identification and authenticating that information for accounting purposes (e.g. RFC 532).

Accounting-type arguments were applied in situations that did not actually involve financial transactions, such as the use of no-cost email systems (RFC 491). Indeed, once accounting entered the conversation, it took over. Some participants found it necessary to point out that this was not the *only* reason to require user identification (RFC 555).

Professionalism. During the first decade of the internet RFCs, professionalism entered the privacy conversation as a norm and as a practice. Many professions require keeping information confidential, and doing so is a mark of professionalism even when confidentiality is not absolutely required. Computing facilities often required assurances that the network