Privacy for Networked Computing, 1969-1979

Sandra Braman
University of Wisconsin-Milwaukee
braman@uwm.edu

"If you have a secret, don't keep it on the ARPAnet."
Brian Harvey, 1975, RFC 686, p. 3.

Those involved in designing what we now refer to as the Internet (initially known as ARPANet[1]) during the first decade of that process (1969-1979) were the first to appreciate what has now become common knowledge: it is very difficult to protect privacy online. This did not mean, though, that the need to try to provide such protections wasn't appreciated from the start. Indeed, about 17% of the 685 documents published through the close of 1979 in the technical document series that records the network design decision-making process -- the Internet Requests for Comments, or RFCs -- included attention to privacy.[2] This is significantly more attention than was given to any other policy issue during that period within the network design discourse.

The Internet is far from the first communication technology to generate privacy concerns. History is filled with other examples. During the 19th and early 20th centuries, what were believed to be 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, eroded class and cultural boundaries that marked the line between that which is public and that which is private by allowing individuals into the home who would not have been permitted to enter previously (Marvin 1988). Between 1930 and 1950, the motorization of the police force and use of the radio in the US significantly diminished the privacy of citizens (Dandeker 1990), and so on.

Protecting the privacy of networked communications was considered both important and problematic by communication network policy-makers beginning with the telegraph, appearing as one of the first topics addressed by the organization formed in 1865 that ultimately became the International Telecommunications Unio (ITU) (Codding, 1972). Early studies of computing (1950-1969), too, included privacy on its list of social problems exacerbated by the technology (Kling, 1980), and privacy was among the topics covered in a

[1] The name "Internet" to refer to the network being built with ARPA support came into use in 1974 (RFC 675).

[2] The final document in 1979 is RFC 758, but a number of documents were not publicly published for either national security or intellectual property rights reasons.

series of articles on the information society in the most prestigious economic newspaper in Japan in 1969 (Ito, 1991).

Early decisions by those who designed the Internet created a situation that enabled the data mining of so much concern to those who would protect civil liberties in the 21st century (Blumenthal & Clark 2001). This paper presents an analysis of the treatment of privacy issues by the computer scientists and electrical engineers involved in the first decade of the design process, 1969-1979, as revealed in the Internet RFCs. This research is part of a larger project analyzing the treatment of legal and policy issues in the RFCs through 1979.[3] The policy frames and the design criteria that served as policy principles developed by Internet designers during the first decade are discussed elsewhere (Braman 2010a). This paper begins with arguments for and against protecting privacy that appeared in the design discussion of the 1970s before going on to look at the range of privacy techniques developed during the first decade and thinking through their long-run effects.

Privacy in the 1970s

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 2010c). For resolving contemporary privacy issues, therefore, laws from governments around the world and from international organizations are pertinent. It was US law, though, that provided the most significant legal context for ARPANet/Internet designers of the 1970s. General perceptions of privacy issues within the Internet design process, and the rapidity with which certain privacy protection practices became the norm, are elements of the design context that help us understand specific arguments and techniques that were put forward.

The Legal Context

Alarm about invasions of personal privacy made possible by government census and labor statistics databases stimulated a series of studies supported by private foundations and the US government that resulted in influential reports during the

[3] An overview of the larger project can be found in Braman (2010b). This material is based on work supported by the National Science Foundation under Grant No. 0823265. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author. Thanks to networking experts David Stack and David Crass of the University of Wisconsin-Milwaukee's University Information Technology Services for their assistance in translating the technical language of the RFCs into terms understandable by the layperson, and to Alyse Below and Liza Barry-Kessler for their research assistance.

19790s (HEW 1973; Privacy Protection Study Commission 1977; Westin & Baker 1972). All of these warned that the significant threats to privacy from computerized databases would be exacerbated when such databases became networked. A book by constitutional scholar Arthur Miller (1971) and a series of books by Alan Westin that began in 1970 brought the issue to the public at large. Regan (2008) and Trubow (1989) provide detailed histories of other developments during this period that included, notably, passage of the Privacy Act in 1974 and the first development of principles for fair information practices to be followed in the digital environment.

Such concerns triggered the attention of social scientists. Since the 1970s, empirical research into the sociology of privacy and the development of theories of privacy have become ever-more important scholarly and research enterprises. We now understand that privacy problems are inherently political (Branscomb 1986; Star & Ruhleder 1996), involving the very boundary-defining activities (Petronio, 2002) that are so flexible -- and thus so complex -- in the networking environment. Often, the same user has competing privacy interests (Case 2000); determining which interest dominates in any given circumstance is a contextual exercise (Nissenbaum 2004). Privacy is typically an issue that suffers from "policy drift" characterized by a lack of an organized constituency and a concomitant reliance on habits developed through the course of practice rather than analysis, debate, and explicit decision-making (Smith 1994). Economically, the costs of invasions of privacy are difficult to quantify -- but not long after the period being reported upon in this study, privacy protection was already being seen as something that could be unbundled from other networked services to be sold separately for profit (Auerbach 1983). Technological innovation introduces new vulnerabilities that are often difficult to foresee because of unfamiliarity or because they result from such complex interactions that they may be unknowable until they occur. All of these perspectives on privacy are evident in the Internet design discourse, though in technical rather than sociologically theoretical terms.

The Network Design Context

Before the design process began, contributors to that process such as the RAND Corporation were already studying the privacy issues that could arise in a digital network environment (RFC 243). Internet designers recognized that there is a difference between reaching a consensus on general principles, such as the importance of protecting privacy, and reaching a consensus on the actual techniques to be used. There is also a tension between establishing standards for privacy protection and the need to minimize constraints on further experimentation and innovation (RFC 195). It was considered difficult to reach agreement on privacy issues because they rarely stand alone. They are often intertwined with other values (eg, security) and issues (eg, accounting). Privacy protection techniques can serve additional technical and social functions as well (RFC 269).

Privacy appeared almost immediately in the Internet RFCs discourse, first in a description of varieties of practice across networked hosts (computer sites) (RFC 109). By 1972, the need for privacy protections at log-in, at least, 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 a masking function on the screen for log-in information (RFC 318). As a categorization system for protocols developed, first privacy (RFC 750), and then security (RFC 753), became running topics. The issue was an inevitable concomitant to discussions about private 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 for the purpose of providing support to remote users (RFC 364). By 1978, testing of sites included an effort to see whether or not it would be possible to send mail to unknown users (RFC 751) -- a practice now called spam and often experienced as an invasion of privacy, when it is undertaken for profit, rather than experimental, purposes.

Throughout the decade, privacy-related concepts underwent further articulation (eg, RFC 435), and descriptions of privacy vulnerabilities appear (eg, RFC 666). Still, some felt that though the issues had been raised, there was insufficient attention to them and that they were constantly being put off for "later" during the design process. Different types of privacy issues were being conflated in a way that wasn't useful (RFC 501). Some authors argued that privacy problems were far worse than was being generally acknowledged (RFC 602). 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 three different perspectives -- that of the network as a whole, that of individual hosts, and that of the user. Distinctions among the three level were evident to network designers (see, eg, RFC 610).

Privacy at the Network Level

Four different types of arguments for the protection of privacy at the network level emerged during the first decade of discussions about Internet design. There was appreciation of the critical role of privacy for network integrity, as an affordance for resource sharing through the network, as a support for accounting systems, and as an element of professionalization.

Protection of Network Integrity. The need to protect network integrity, which requires both security and privacy, was so important that the ultimate abandonment of one proposed protocol in which there had been a lot of interest (RJE) was attributed to its weaknesses on this front (RFC 725). An early expression of the general need to ensure network integrity (RFC 62) became unbundled into a number of different elements as the first decade of the design process progressed. The importance of creating an environment of trust, widely recognized as fundamental to the success of any type of networked activity in the 21st century, was first explicitly articulated by network designers in 1971 (RFC 98). Users of protected file systems, it was argued, should be able to have a reasonable degree of confidence that servers they are using are able to identify remote users correctly (RFC 114). A trickle-down argument was made that developing privacy protections to military specifications would result in enhanced privacy for all network users (RFC 316).

As is the case throughout the Internet design process, humans as users and computing processes as users ("daemons") required separate attention (Braman 2010c). Although we aren't accustomed to applying the concept of privacy to non-humans, server processes, too, needed to be able to securely exchange socket names in order to establish a trusted connection (RFC 430). For the purpose of determining access rights, it was unclear just which identity/identifier should apply to computing processes (RFC 501), or how a server should determine whether or not any given process required a distinct log-in process involving verification of identity (RFC 555). In another parallel between daemon and human users, the identity of a sending or receiving computer socket was considered information that needed privacy protection under some circumstances (RFC 54).

Enabling Resource Sharing. Resource sharing was defined as a form of interprocess communication that linked specific resources to particular processes (RFC 61), an orientation that frames resource sharing issues from the daemonic rather than human perspective. However, very early on it was recognized that human user privacy was essential to what they then referred to as "indirect" use of networked computers (RFC 114) -- that is, computing distributed beyond a single machine and/or at a distance. Network designers quickly became aware that as use of database systems in the network grew, so would the urgency of the need for privacy protections (RFC 340), even when private connections continued to be used for batch processing by some users (RFC 647). Privacy was one of six broad areas identified as crucial for data sharing in 1971 (RFC 146), equal in importance to the abilities to manipulate files across systems, computerize databases, logically restructure data in response to queries while holding the physical structure stable, use data management systems without needing to engage with the specifics of processing on any given computer, and keep duplicate copies of a database consistent.

Accounting. As soon as ARPA-funded host sites opened themselves up for experimentation by users without ARPA support, the question of keeping track of the use of network and computational resources for accounting purposes appeared. It was quickly understood that authentication of user identities was necessary (RFC 136). Billing implications of access controls for the network arose when retrieving files from one computer for use on another as well as when using the computational capacities of a computer other than one's own. If those using network resources were not identified, the cost of providing services would necessarily become part of the system overhead for the serving host (RFC 487).

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). Once accounting had entered the conversation, some participants found it necessary to remind others that this was not the *only* reason to require user identification; preventing fraud, and what we would now call spoofing, were important reasons as well (RFC 555). Passwords were an obvious means of both requiring user identification and authenticating that information for accounting purposes (eg, RFC 532).

Professionalization. Professional dimensions of privacy were evident in the first decade of Internet Requests for Comments documents in two different ways -- it was discussed as a norm and demonstrated as a practice. Incorporating privacy protections into the network was understood to a normative requirement for the kind of professionalization that would be needed in order to support the connection of additional computers to the network (RFC 111). Then, as now, there were suggestions that those who did data entry should be licensed in order to be able to monitor their integrity, accuracy, and accountability. Keeping private information confidential is a behavioral requirement for many professions, and a mark of professionalism in others in which it is not absolutely required. The wider practice of keeping comments about specific individuals anonymous and thus the identity of those being discussed confidential was modeled in the RFCs when an author reporting on reliability issues refrained from naming a particular site that was being experienced as strikingly unreliable (RFC 282).

Privacy at the Host Level

Although the phrase is not used either in the privacy literature in general or within the RFCs when privacy is discussed, what can be described as the network externalities of privacy for networked hosts were recognized during the first decade of the design process. The externalities considered important enough to take into design consideration during this period included protecting host integrity, the necessity of user verification, and the value of enabling privacy during experimental use. There were significant differences across hosts in terms of the level of attention paid to privacy matters and the types of techniques used (RFC 109).

Protecting Host Integrity. For the host, it was understood that protecting the privacy of specific content was the only way to ensure the protection of all content and of the system itself. Permitting even one inauthentic user to access files on its system would place all stored files and data at risk. Thus protecting privacy was a matter of protecting the integrity -- and the reputation -- of the host itself (RFC 49).

User Verification. The idea that serving hosts should require users to identify themselves through the use of user names and passwords, at minimum, is a notion that appeared early in the decade and appeared repeatedly throughout. A spectrum of levels of detail and of types of information required for this purpose was acknowledged, depending on what it was that was being protected (RFC 163). In some cases, access controls were needed at the file level; in others, it was also needed at the level of data within a file (RFC 164); and for yet others – the military -- verification also had to take place in order to access a networked terminal (RFC 316). Having access to data was distinguished from the right to modify data (RFC 269), and the same distinction was drawn for passwords (RFC 463). The introduction of satellite linkages, which entailed long time delays during the 1970s, created an additional user verification problem (RFC 357).

Passwords and account information are understood to be sensitive information the privacy of which must also be protected (RFC 385). For those receiving information, authenticating the identity of someone from whom a message or request is being received is necessary in order to have confidence that the sender is actually the user it is claimed to be. The level of trust in any verification mechanism, in turn, depends upon the level of confidence that the source host's user authentication and access control mechanisms are accurate (RFC 644).

Because there are different motivations for protecting diverse types of data, distinctions among types of users were also important for network designers during the first decade. It can be important to distinguish between sponsors of data and those who use the information (RFC 144), or to separately require user identification at the stages of input and output (RFC 360). In many cases, users need not be identifiable at the individual level but, rather, at the group level for privacy purposes; three options were to grant access to a specific individual, to members of an identifiable group, to everyone who has already been granted log-on privileges to a certain computer, and to the public at large (RFC 487). Though most systems allowed users to choose their own passwords, at least one institution was assigning user id-password pairs during the 1970s (RFC 436).

Privacy at the User Level

The simple fact that many users prefer privacy in computing networks is its own justification for incorporating such protections into the network (RFC 90). That preference derives from a range of concerns about who has access to, and who can manipulate, content of various types. There is the need to protect particular files from unauthorized or accidental use (RFC 114). There are national security concerns, whether for data (RFC 90), files (RFC 316), or voice (RFC 741). The level and nature of user preference for privacy protections can vary with the type of data involved (RFC 144). Network designers acknowledged differences in such requirements as applied to medical, criminal justice, and transactional data. Social security numbers were one example of a type of non-password information for which individual users would be keenly interested in privacy protection (RFC 731). Invasions of data privacy were linked with threats to data integrity (RFC 98) as well as to the quality of data representations and what we now refer to as metadata and information architecture (RFC 327).

User preferences for privacy protection were expressed not only as they arose for material as it is stored but also during transit, whether that material involves files (RFC 354) or communications between human users (what we now call email) or daemon users (RFC 524). Privacy protections were most often conceived of as access control mechanisms, keeping entities from accessing content to which they don't have rights. As was noted in RFC 49, though, techniques for protecting privacy are also a means of ensuring access itself to rightful users, for malicious users can make it impossible for others to get in.

Three drivers of user preference for privacy protections were particularly human in nature. First, the desire to protect secrets was acknowledged (RFC 318). Second, reflecting a US history in which the right to anonymous speech is constitutionally protected, there was also respect for the secrecy of authorship (eg, RFC 282); it was expected that there would be anonymous users (eg, RFC 450), and RFC 549 was authored anonymously. And sometimes the argument was overtly political; one author declared, "I'm afraid that I can't work up much excitement about helping the CIA keep track of what anti-war demonstrations I attended in 1968 . . . ." (RFC 686, p. 1).

Arguments against Protecting Privacy

A number of reasons for *not* designing privacy protections into the network were also presented during the first decade of the design discussion. There were arguments arguments based on utopian and/or political perspectives as well as those that derived from placing system efficiency at the top of the hierarchy of values being pursued. Vulnerabilities introduced by privacy protections are not arguments against privacy *per se*, but they might be used as such in some circumstances.

Utopian/Political Arguments

Though larger claims about the utopianism of network designers in their early years have been made (see, eg, Turner 2006), it is also probable that the initial trust among members of the network community derived from the small and intimate nature of that group. Some believed there was no need for privacy protections because all processes launched by system users would be "good" (RFC 62, p. 3), and/or it was sufficient to rely upon the protections provided by the serving host (RFC 114).

A second type of utopian argument emphasized the importance of free access, with email providing a focal example of both a network process (RFC 475) and of content (email that had been "journalized" by the Network Information Center) (eg, RFC 629) that should be available to all anonymously (RFC 694).

Efficiency Arguments

All policy-making involves trade-offs among multiple values of social importance. Network designers who valued efficiency above all else expressed concerned during the 19790s that privacy protections would impede their ability to achieve system efficiency (RFC 172). They wanted daemon users (computer processes) to be able to move in a fluid manner among users (RFC 61), and human users to have easy access to publicly available files (RFC 487).

Privacy protections did introduce constraints that made the design job more complicated. Jon Postel argued that the goal of finding a way to mask input in order to protect privacy should be dropped because it was too difficult -- it was impossible to know just how much input to mask because passwords and other secure information are of variable length (RFC 328). Other designers quickly pointed out that just because a task was hard didn't mean it couldn't -- or shouldn't -- be done (RFC 340). The solution to this problem developed over the course of a multiple-document conversation through which a consensus was ultimately reached on a still-familiar solution: systems can specify either the exact number, or minima and maxima, of characters to be used in a username or password as well as in file names and directory pathnames (RFC 607).

For users, the efficiency of data sharing is necessarily reduced with encryption, since keys would be shared by a pair of communicating individuals only (RFC 753). Although a 1971 evaluation of responses to the use of password protection at one university showed that users found passwords easy to use (RFC 269), some Internet designers that users would get tired of having to type in their usernames and passwords all the time (RFC 491).

Vulnerability Arguments

Internet designers quickly learned that privacy protections can introduce vulnerabilities to both privacy and to protocols themselves. These weaknesses could be either human or technical. At the intersection of the two are those matters that were treated as human errors during the first decade of the design process, but solutions to which were ultimately incorporated into software, becoming technical matters.

There was evidence that such vulnerabilities did allow the network to be hacked. In 1973 it was reported that at least two major serving hosts crashed under suspicious circumstances by individuals who should have known what they were risking. On a third system, the method of establishing passwords was compromised by two high school students (RFC 602). Since experimentation with hacking -- "phreaking" -- of the telephone network had begun in 1957, with the introduction of automatic switches, it should not be surprising that there was an active subculture ready to work on breaking into the new packet switching network.

Human Failures. A number of types of what we now popularly refer to as "operator errors" that defeat or undermine privacy protection efforts were mentioned in RFCs from 1969 through 1979. Some of these are still familiar today, while others derived from the administrative systems of the time.

Individual sites accustomed to relying on physical isolation for protection didn't immediately recognize that new procedures had to be used in a networked environment. People commonly used passwords that were easy to guess. The telephone numbers of host sites were published far more widely than intended, or than most understood; one author likened their distribution to that of phone numbers on walls of phone booths or men's rooms (RFC 602). Inconsistent use of names, nicknames, and initials in addressing was problematic (RFC 757).

A study of the use of ARPANet nodes by University of California-Santa Barbara students in 1972 found that it was frustrating for users when passwords for those with free use were randomly changed without alerting users to the fact (RFC 369). The Network Information Center (NIC) that was providing administrative support to the networking effort maintained a list of all of those on the network, but this list was inaccurate, left out nicknames in common usage, and was designed in such a way that it was difficult to put into computer memory (RFC 752). Some found it difficult to navigate differences in editing systems when trying to identify themselves and to verify the identity of others (RFC 475).

Some mistakes were amusing. One RFC author reports that a particular ARPANet host was so suspicious of those not at the local site that it randomly generated a new password every week for the use by those at other sites -- and then sent the new password to those users through unprotected email. Those who received the email typically copied the information into an unprotected file on their hard drives for ease of use, so not one but two vulnerabilities were introduced into the system by this purported effort to protect privacy and security (RFC 686).

Technical Failures. The state of network design during the 1970s left openings for several types of invasions of privacy. At the simplest level, a number of computers simply did not "respect," or make use of, network IDs generated by the Network Information Center for identity verification purposes (RFC 475). At a second level, users were able to game the system to serve their own purposes. There were several techniques by which an individual who had not been granted the privilege could gain access to the protected files of someone else (RFC 505), including use of a process for having files mailed through the network (RFC 487).

A third type of technical vulnerabilities arose from complexities of interactions among diverse elements of the network. A system put in place to test changes to such fundamental elements of the system as the computer core and computer code loaded to enable connections to the network did not initially include a means of analyzing unauthorized activity. It was recognized by 1973, however, that this would be necessary to protect against what were already being referred to as "hackers," though it was also acknowledged that this would not be sufficient to protect against "a determined and malicious attack" (RFC 521, p. 2). An unintended consequence of the name/finger program, which allows remote users to see a "friendly, human-oriented status report" about who is using a given host, is that it provides so much information about those users that it could well be experienced as invading a user's privacy (RFC 752).

Failures at the Human/Daemon Intersection. Two problems with privacy implications arose during 1969-1979 that were perceived to derive from human error at the time but that later were addressed through programming and treated as a technical matter. Both involved keeping databases current and correct when data within them changes. Data changes can require simultaneous alterations to index information (RFC 219), and obsolete information needs to be removed from databases (RFC 677).

Techniques for Protecting Network Privacy

It was recognized early on that a wide range of types of privacy protections was available; the poles of a spectrum of approaches distinguished by degree of complexity as described in 1971 went from knowledge of the pathname to a particular file, with password protection for the directory, to an elaborate hierarchy of group-project-task-username membership with separate controls for reading and writing (RFC 180). Access controls were to include specification of whether more than one user can simultaneously be updating a file; whether a file creator can specify authorized users and, if so, how; and whether or not it was possible to put in place different access controls for different subunits of a given file (*Ibid*.) Access controls were defined as a means of defining users' access privileges to the use of a system and to files in that system (RFC 354).

The myriad techniques for privacy protection that were proposed and/ or underwent experimentation during the period 1969-1979 can be categorized according to whether they were methods that would be used by humans, by network processes, or by those working with data. There was extensive discussion in the RFCs during the first decade regarding just where responsibility for protecting privacy should belong and the need to disperse techniques throughout the infrastructure, but there is insufficient space here to cover these issues.

Human Techniques

Using identification information at the point of logging in received the most attention is a means of protecting privacy during the first decade of the Internet design process. Other approaches, though, were also mentioned, including making agreements offline and masking input. Many systems set up for local users of a site, where all users had personal knowledge of those who had access and informal procedures sufficed, didn't work or weren't available for remote users (RFC 364) – exemplifying telecomunications policy analyst Noam's (1992) insight that privacy is an example of an intraorganizational issue that becomes a public matter once an organization is networked.

Logging In. Use of a password at the point of logging in to a specific server was first mentioned in RFC 48, when several familiar options were mentioned: allowing everyone on, requiring a recognized identifier (which could be a user name), requiring a password, or requiring both an id and a password. The log-in dialog was conceived of as two-fold, involving both what to say and how to say it (RFC 98). During the early years, the password would actually not be accepted until the receiving system knew that it had sufficient space to accommodate another user (RFC 122). It wasn't long before account numbers also came into use as an additional identifier (RFC 223). Many systems that permitted anyone to become a user, at least to learn what capabilities were offered by the host, still required log-on information but provided a common identifier for all to use (RFC 265). Though at one point it was believed that systems did not need to respond to receipt of the information (*Ibid.*), reply codes did soon come into use to report on the success or failure of the communication and/or the connection itself (RFC 640).

With time, log-in processes became more elaborate. The log-in detail included in FTP included attention to such matters as flushing identifier information from the system after use and masking the input (RFC 542), techniques discussed further below. It became clear that servers need to verify identifier information provided by users, a function provided at the time by the NIC (RFC 555), but that many believed would be better provided by another third party service (eg, RFC 462). Some hosts required separate identifiers for specific tasks once on a system (RFC 360), or specified that a given set of log-in identifiers could only be used by a single user at a time (RFC 477).

Though many of the ideas from the 1970s about how to handle log-in practices are still in use, others came and went. The counterintuitive practice of submitting user identifier information at any point during a session rather than at the beginning was permitted for a while (RFC 265), but soon went out of use. One author proposed the concept of a network "birthplace," the site from which a user first comes onto the network, as the place at which a unique lifelong network identifier would be generated that would follow the user from site to site (RFC 757).

Masking Input. The echoplex function, which first came into use for log-in purposes, sends typed material directly to the computer and the computer echoes it back to the printer (RFC 98). It was quickly realized that passwords shouldn't be readable if they are to provide privacy protection, so the "hide your input" command directed a printer to suppress printing (RFC 158). Ultimately the echoplex and hide your input functions were separated, with the latter understood as a special case of the former (RFC 393). Not all hosts during the 1970s were able to support the echo function and concomitant ability to mask input, though (RFC 393), so a description of a proposed directory service included information about whether each host expects a terminal to echo locally or remotely (RFC 608).

It was because of the need to support the masking of input that limits were put on the length of passwords; the initial recommendation was that passwords should be limited to 8 characters, and user names to 32 (RFC 614). The dysfunctionality of masking all mail content of mail was an argument against sending network mail directly to printers because of the lack of privacy (RFC 475).

Offline Arrangements. One of the earliest privacy protection techniques discussed in the Internet RFCs was the very human approach of establishing a connection only after previously agreeing to do so by telephone or letter. This was thought to be a good means of addressing the problem of "how are both users to be confident that they are talking [with] each other, and not some interloper?" (RFC 129, p. 2). The authors of this 1971 document were leery of the directory approach for verifying identity because they believed it would make computers more vulnerable to attack.

Network-Based Techniques

One of the first expressions of the sense that the network protocols should be law-like in nature came in response to the problem of identifying users:

> it should be a basic protocol law that *no process whatsoever* may request or accept connections or transmit or receive data over a socket having a user code not its own (RFC 49, emphasis in the original, p. 4).

Four approaches to protecting privacy at the network level developed during the 1970s: keeping some aspects of networking private, termination of activity, using elements of message design for this purpose, and establishing connection identities.

Private Networking. Although the goal of the ARPANET project was to build a network for widespread use, it was understood from the start that -- at least for some purposes, and for some users -- there are times when it would be desirable to cordon off networking activity. The concept in 1970 was that doing so would create a subset that would connect with the larger network but be separated from it (RFC 54).

Off-line storage, using privately owned disk packs, also came into use early as a means of protecting privacy (RFC 90). Documentation about the network itself that needed to be kept private (for either national security or intellectual property rights reasons) was sent to those who should receive it individually as a memo rather than using the NIC mail service that distributed documentation to all (RFC 82).

Termination of Activity. Terminating activity came into use as a network privacy protection mechanism that could affect either processes or content. As soon as an error message was received from a remote host, it was argued, the serving host should shut down all processes to protect both local data and remote user privacy (RFC 98). Additional triggers for terminating activity as a privacy protection were identified with experience, including closing the connection if the user name and password aren't completed within a specified time period (RFC 360). In a variation on the theme, one facility experimented with closing the connection used for log-in purposes and opening a second one for transmitting files (RFC 310).

The reinitialize command in FTP terminates a user, flushing all input and output information as well as account information (RFC 454). This clears buffers, but also provides some privacy protection and thus the idea came to be taken up by hosts that would flush user names and associated passwords from their systems once a user has no jobs on that system (RFC 477). Harvard University went so far as to delete all files associated with a terminal that was no longer active (RFC 499), a practice that would have been quite problematic for the networked computing effort had it been sustained or widespread since it seemed to use a relatively short time horizon for such a decision.

Message Design. Two features of message design were useful from a privacy perspective: packetizing content, and the structure of headers. The ARPANET project, and the Internet today, are packet-switched networks (see, eg, RFC 675). In the traditional wired telegraphy and telephony environment, messages were moved around the network through line switching, in which a physical line connects two pieces of equipment and messages or conversations in their entireties are moved from one line to another using either the manual switching equipment of a telephone or telegraph switchboard, or electronically switched. In the packet switching environment, messages are broken up into many packets, each with its own header, and each with its own path to the receiver, with the whole being reassembled into a coherent message only upon receipt. Packetizing in itself provides some privacy protection for content while it is in transmission when only fragments of a message are intercepted along a given path.

Several ideas were put forward in the 1970s about information to include in message headers that had privacy implications. Inclusion of an "authentication" field provides information about which originator fields have been authenticated, and by which systems, and the "BCC" field was believed to be useful for access control (RFC 680). It was suggested that the header should tell users whether or not the connection in use is secure (RFC 717). One proposal not acted upon was to include an "FCC" field in the header that would tell users where messages were being stored (RFC 724).

An interesting feature of the header discussion as it pertained to privacy was the line between providing information of use to humans versus providing information of use to daemons. The initial proposal to include authentication information in the header confessed that "This document attempts to tread the narrow line between features for human processing and features for machine processing" (RFC 680, p. 1). The fields listed were meant to be useful to humans even if automatic processing were not supplied, and instructions within angle brackets were intended to provide machine-readable information regarding the need of a daemon to look at any particular field. Still, 2 years after publication of RFC 680, it was felt necessary to remind those involved in designing the network that it was necessary to make sure that mail information -- including fields related to privacy and security -- is readable by humans (RFC 724).

<u>Connection Identity</u>. Within the first year of the design process, each computer on the network was given a private subset of unique identifiers for its sockets so that connections made could be named by the pair of sockets linked (RFC 54). These identification numbers provided some assurance that the user asserted was actually the user involved (RFC 61), and it could be set up so that each socket could connect with only one process (RFC 675). Users could ensure the security of data transfer by specifying that connections would be accepted only from specific hosts and sockets (RFC 438). Within the decade, it was found necessary to develop messages to be sent if a security or privacy issue were suspected at the point of connection (RFC 686), and authentication issues began to receive attention (RFC 739).

Data-Based Techniques

Encryption is a well-known privacy protection technique that is accomplished by working with the information being protected rather than the network or the network user. Both the structuring and the labeling of data provide additional opportunities for privacy protection.

Information Architecture. The notion that information architecture -- the ways in which information is structured -- can be used as a privacy protection technique has become a 21st century policy issue with the digitization of health records. Those involved with designing the Internet, though, appreciated the privacy protection potential of information architecture in the 1970s. There was early discussion of the value of establishing pathnames, which locate files within the information architecture of a particular computer, for enabling file-specific access controls such as passwords (RFC 114). The privacy and security value of such information was considered so important that it was suggested that file directories themselves should be restricted access (RFC 219).

What we now call metadata also came into use during the early years of the Internet design process with the design of computers and "data languages" that stored information *about* data separately from the data itself. Access controls could then be oriented around this metadata rather than requiring user specification of controls at the moment of the creation of each individual file (RFC 219).

Encryption. Encryption receives its first mention in the RFCs, other than in a bibliography, as a technical means of protecting privacy that would be complementary to the use of metadata (RFC 610). It was the desire to use the network for voice communications by the military that stimulated interest in encryption (RFC 720). An extensive scheme for encryption was presented as part of the "Internet Message Protocol" (IMP) in 1979. The approach used allowed encrypting messages either as a whole or in part, and the fact that all parts of a message could be encrypted -- including header information -- was specifically mentioned (RFC 753).

Conclusions

A lot can be learned about privacy as a policy issue for those building, using, and regulating the Internet from those involved during the first decade of the design process, 1969-1979. They recognized that privacy is a multi-dimensional problem, that it arises at every stage of networking, and that it has to be revisited every time there is a change in technologies. They understood that the same user may hold conflicting views on privacy, depending on which activity is being undertaken and the role held. And they knew that the introduction of one technique for protecting privacy could open up other possible means of invading privacy.

Network designers during the 1970s appear extremely sophisticated in their thinking about privacy when evaluated vis-a-vis theoretical developments since that time. They viewed privacy as contextual and well knew that it involves boundary-setting. They were clear-eyed regarding tensions between privacy and the achievement of other goals such as national security and efficiency. Information architecture and metadata were used by these electrical engineers and computer scientists as tools for privacy protection.

Future work will continue to trace the development of thinking about and techniques for protecting privacy on the Internet as it moved forward into the present. For now, policy-makers can take away the message that general statements about protecting data privacy are inadequate. To protect privacy in the digital network environment, legal and regulatory mandates must be more specific in detailing the various sites and processes at which or during which privacy must be protected. For mandates regarding privacy protection techniques to make sense, law-makers should be working together with those in the technical community rather than in isolation or at contrapoint.

References

Auerbach, L. 1983. "Privacy and Canadian Telecommunications Regulation." *Telecommunications Policy 7*(1): 35-42.

Blumenthal, M. S. & Clark, D. D. 2001. "Rethinking the Design of the Internet: The End-to-End Argument vs. the Brave New World. *Transactions on Internet Technology 1*(1): 70-109.

Braman, S. 2010a. "The Framing Years: Policy Fundamentals in the Internet Design Process." To be presented to the Telecommunications Policy Research Conference, October, Arlington, VA.

Braman, S. 2010b. "Internationalization of the Internet by Design: The first Decade." Unpublished manuscript.

Braman, S. 2010c. "The Interpenetration of Technical and Legal Decision-Making for the Internet." *Information, Communication & Society 13*(3): 309-324.

Branscomb, A. W. 1986. *Toward a Law of Global Communication Networks*. New York: Longman.

Case, D. O. 2000. "Stalking, Monitoring, and Profiling: A Typology and Case Studies of Harmful Uses of Caller ID." *New Media & Society 2*(1): 67-84.

Codding, G. A. Jr. 1972. *The International Telecommunications Union: An Experiment in Iinternational Cooperation*. New York: Arno Press.

Dandeker, C. 1990. *Surveillance, Power and Modernity: Bureaucracy and Discipline from 1700 to the Present Day*. New York: St. Martin's Press.

Ito, Y. 1991. "*Johoka* as a Driving Force of Social Change." *KEIO Communication Review 12*: 33-58.

Kling, R. 1980. "Social Analyses of Computing: Theoretical Perspectives in Recent Empirical Research." *Computing Surveys 12*(1): 61-110.

Marvin, C. 1988. *When Old Technologies Were New: Thinking about Electronic Communication in the Late Eighteenth Century*. New York: Oxford University Press.

Miller, A. 1971. *The Assault on Privacy: Computers, Data Banks, and Dossiers*. New York: Signet.

Nerone, J. 1994. *Violence against the Press: Policing the Public Sphere in US History*. New York: Oxford University Press.

Nissenbaum, H. 2004. "Privacy as Contextual Integrity." *Washington Law Review 79*(1): 119-158.

Petronio, S. S. 2002. *Boundaries of Privacy: Dialectics of Disclosure*. Albany, NY: State University of New York Press.

Privacy Protection Study Commission. 1977. *Personal Privacy in an Information Society*. Washington, DC: Government Printing Office.

Regan, P. M. 2008. "The United States." In J. B. Rule and G. Greenleaf (Eds.), *Global Privacy Protection: The First Generation*, pp. 50-79. Edward Elgar Publishing.

Smith, H. J. 1994. *Managing Privacy: Information Technology and Corporate America*. Chapel Hill, NC: University of North Carolina Press.

Star, S. L. and Ruhleder, K. 1996. "Steps toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research 7*(1): 111-134.

Trubow, G. 1989. *Watching the Watchers: The Coordination of Federal Privacy Policy*. Washington, DC: Benton Foundation.

Turner, F. 2006. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago: University of Chicago Press.

US Department of Health, Education, and Welfare (HEW). 1973. *Computers, Records, and the Rights of Citizens*. Washington, DC: US Department of Health, Education, and Welfare.

Westin, A. F. 1970. *Privacy and Freedom*. New York: Atheneum.

Westin, A. F. and Baker, M. A. 1972. *Data Banks in a Free Society*. New York: Times Books.

RFCs Cited

RFC 48. *Possible Protocol Plateau*. J. Postel, S. Crocker. April 1970.
RFC 49. *Conversations with S. Crocker (UCLA).* E. Meyer. April 1970.
RFC 54. *Official Protocol Proffering*. S. D. Crocker, J. Postel, J. Newkirk, M. Kraley. June 1970.
RFC 61. *Note on Interprocess Communication in a Resource Sharing Computer Network*. D. C. Walden. August 1970.
RFC 62. *Systems for Interprocess Communication in a Resource Sharing Computer Network*. D. C. Walden. July 1970.
RFC 82. *Network Meeting Notes*. E. Meyer. December 1970.
RFC 90. *CCN as a Network Service Center*. R. T. Braden. January 1971.
RFC 98. *Logger Protocol Proposal*. E. Meyer, T. Skinner. February 1971.
RFC 109. *Level III Server Protocol for the Lincoln Laboratory 360/67 Host*. J. Winett. March 1971.
RFC 111. *Pressure from the Chairman*. S. D. Crocker. March 1971.
RFC 114. *File Transfer Protocol*. A. K. Bhushan. April 1971.
RFC 122. *Network Specifications for UCSB's Simple-Minded File System*. J. E. White. April 1971.
RFC 129. *Request for Comments on Socket Name Structure*. E. Harslem, J. Heafner, E. Meyer. April 1971.
RFC 136. *Host Accounting and Administrative Procedures* R. E. Kahn. April 1971.
RFC 144. *Data Sharing on Computer Networks*. A. Shoshani. April 1971.
RFC 146. *Views on Issues Relevant to Data Sharing on Computer Networks*. P. M. Karp, D. B. McKay, D. C. M. Wood. May 1971.
RFC 163. *Data Transfer Protocols*. V. G. Cerf. May 1971.
RFC 164. *Minutes of Network Working Group Meeting, 5/16 through 5/19/71*. J. F. Heafner. May 1971.
RFC 172. *The File Transfer Protocol*. A. Bhushan, B. Braden, W. Crowther, E. Harslem, J. Heafner, A. McKenzie, J. Melvin, B. Sundberg, D. Watson, J. White. June 1971.
RFC 180. *File System Questionnaire*. A. M. McKenzie. June 1971.
RFC 195. *Data Computers: Data Descriptions and Access Language*. G. H. Mealy. July 1971.
RFC 219. *User's View of the Datacomputer*. R. Winter. September 1971.
RFC 223. *Network Information Center Schedule for Network Users*. J. T. Melvin, R. W. Watson. September 1971.
RFC 243. *Network and Data Sharing Bibliography*. A. P. Mullery. October 1971.

RFC 265. *The File Transfer Protocol*. A. Bhushan, B. Braden, W. Crowther, E. Harslem, J.Heafner, A. McKenzie, J. Melvin, B. Sundberg, D. Watson, J. White. November 1971.

RFC 269. *Some Experience with File Transfer*. H. Brodie. December 1971.

RFC 282. *Graphics Meeting Report*. M. A. Padlipsky. December 1971.

RFC 307. *Using Network Remote Job Entry*. E. Harslem. February 1972.

RFC 310. *Another Look at Data and File Transfer Protocols*. A. K. Bhushan. April 1972.

RFC 316. *ARPA Network Data Management Working Group*. D. B. McKay, A. P. Mullery. May 1972.

RFC 318. *Telnet Protocols*. J. Postel. April 1972.

RFC 327. *Data and File Transfer Workshop Notes.* A. K. Bhushan. April 1972.

RFC 328. *Suggested Telnet Protocol Changes* J. Postel. April 1972.

RFC 340. *Proposed Telnet Changes*. T. C. O'Sullivan. May 1972.

RFC 354. *File Transfer Protocol*. A. K. Bhushan. July 1972.

RFC 357. *Echoing Strategy for Satellite Links*. J. Davidson. June 1972.

RFC 360. *Proposed Remote Job Entry Protocol*. C. Holland. June 1972.

RFC 364. *Serving Remote Users on the ARPANET*. M. D. Abrams. July 1972.

RFC 369. *Evaluation of ARPANET Services January-March, 1972*. J. R. Pickens. July 1972.

RFC 393. *Comments on Telnet Protocol Changes*. J. M. Winett. October 1972.

RFC 385. *Comments on the File Transfer Protocol*. A. K. Bhushan. August 1972.

RFC 430. *Comments on File Transfer Protocol*. R. T. Braden. February 1973.

RFC 435. *Telnet Issues*. B. Cosell, D. C. Walden. January 1973.

RFC 436. *Announcement of RJS at UCSB.* M. Krilanovich. January 1973.

RFC 438. *FTP Server-Server Interaction*. R. Thomas, R. Clements. January 1973.

RFC 450. *MULTICS Sampling Timeout Change*. M. A. Padlipsky. February 1973.

RFC 454. *File Transfer Protocol: Meeting Announcement and a New Proposed Document*. A. M. McKenzie. February 1973.

RFC 462. *Responding to User Needs*. J. Iseli, D. Crocker. February 1973.

RFC 463. *FTP Comments and Response to RFC 430*. A. K. Bhushan. February 1973.

RFC 475. *FTP and Network Mail System*. A. K. Bhushan. March 1973.

RFC 477. *Remote Job Service at UCSB.* M. Krilanovich. May 1973.

RFC 487. *Free File Transfer*. R. D. Bressler. April 1973.

RFC 491. *What is "Free"?* M. A. Padlipsky. April 1973.

RFC 499. *Harvard's Network RJE*. B. R. Reussow. April 1973.

RFC 501. *Un-muddling "Free File Transfer"*. K. T. Pogran. May 1973.

RFC 505. *Two Solutions to a File Transfer Access Problem*. M. A. Padlipsky. June 1973.

RFC 521. *Restricted Use of IMP DDT*. A. M. McKenzie. May 1973.

RFC 524. *Proposed Mail Protocol*. J. E. White. June 1973.

RFC 532. *UCSD-CC Server-FTP Facility*. R. G. Merryman. July 1973.

RFC 542. *File Transfer Protocol*. N. Neigus. August 1973.

RFC 555. *Responses to Critiques of the Proposed Mail Protocol*. J. E. White. July 1973.

RFC 602. *"The Stockings were Hung by the Chimney with Care."* R. M. Metcalfe. December 1973.

RFC 607. *Comments on the File Transfer Protocol* M. Krilanovich, G. Gregg. January 1974.

RFC 608. *Host Names On-Line*. M. D. Kudlick. January 1974.

RFC 610. *Further Datalanguage Design Concepts*. R. Winter, J. Hill, W. Greiff. December 1973.

RFC 614. *Response to RFC 607: "Comments on the File Transfer Protocol."* K. T. Pogran, N. Neigus. January 1974.

RFC 640. *Revised FTP Reply Codes*. J. Postel. June 1974.

RFC 644. *On the Problem of Signature Authentication for Network Mail*. R. Thomas. July 1974.

RFC 647. *Proposed Protocol for Connecting Network Computers to ARPA-Like Networks via Front-End Processors*. M. A. Padlipsky. November 1974.

RFC 666. *Specification of the Unified User-Level Protocol*. M. A. Padlipsky. November 1974.

RFC 675. *Specification of Internet Transmission Control Program*. V. Cerf, Y. Dalal, C. Sunshine. December 1974.

RFC 677. *Maintenance of Duplicate Databases*. P. R. Johnson, R. Thomas. January 1975.

RFC 680. *Message Transmission Protocol*. T. H. Myer, D. A. Henderson. April 1975.

RFC 686. *Leaving Well Enough Alone*. B. Harvey. May 1975.

RFC 694. *Protocol Information*. J. Postel. June 1975.

RFC 717. *Assigned Network Numbers*. J. Postel. July 1976.

RFC 720. *Address Specification Syntax for Network Mail*. D. Crocker. August 1976.

RFC 724. *Proposed Official Standard for the Format of ARPA Network Messages*. D. Crocker, K. T. Pogran, J. Vittal, D. A. Henderson. May 1977.

RFC 725. *RJE Protocol for a Resource Sharing Network*. J. D. Day, G. R. Grossman. March 1977.

RFC 731. *Telnet Data Entry Option*. J. D. Day. June 1977.

RFC 739. *Assigned Numbers*. J. Postel. November 1977.
RFC 741. *Specifications for the Network Voice Protocol (NVP)*. D. Cohen. November 1977.
RFC 750. *Assigned Numbers*. J. Postel. September 1978.
RFC 751. *Survey of FTP Mail and MLFL*. P. D. Lebling. December 1978.
RFC 752. *Universal Host Table*. M. R. Crispin. January 1979.
RFC 753. *Internet Message Protocol*. J. Postel. March 1979.
RFC 757. *Suggested Solution to the Naming, Addressing, and Delivery Problem for ARPANET Message Systems*. D. P. Deutsch. September 1979.
RFC 1122. *Requirements for Internet Hosts: Communication Layers*. R. Braden, Ed. October 1989.