Resource Discovery and Privacy

Michael F. Schwartz
University of Colorado - Boulder
Published in Internet Society News 2(1), Spring 1993

As resource discovery and information services proliferate on the Internet, a number of privacy issues arise. The most obvious examples concern directories of people. What rights should people have to control what information about themselves is visible, who can access this information, and how the information is updated?

Privacy problems arise in other types of directories as well. For example, from time to time users of the archie FTP directory service [Emtage & Deutsch 1992] discover information that was intended only for limited distribution. Typically this happens when a user looking for a particular program or document happens across other information while browsing the FTP site where the needed information was located. This happened, for example, with an early draft of a document being created by a networking organization last year.

Another privacy problem arises in conjunction with the provision of directory service: access logs (which are routinely collected by many Internet services) can be used to determine peoples' interests, relationships, and other sensitive information. For example, logs of an individual's use of archie or Gopher could indicate interests in particular personal discussions. While it has always been possible to collect such information from older, non-directory oriented services (such as USENET news servers), the current generation of directory services can collect information about a much larger, more widely distributed collection of network users. It is typical for such services to receive requests from thousands of users around the world every day.

In some cases, an explicit directory is not even needed to violate privacy. For example, it is possible to determine shared interest relationships between thousands of people by briefly monitoring and analyzing electronic mail "To:/From:" logs from a handful of sites [Schwartz & Wood 1992]. It is difficult to protect against this sort of invasion, because the needed data are not easily masked. Privacy Enhanced Mail [Linn 1987] does not protect against this type of analysis, because the actual content of mail messages (which is what PEM protects) is not needed for the analysis. Protecting against this type of traffic analysis [Callimahos 1989] would require generating spurious traffic, to hide patterns in the real traffic. Doing so could be costly.

In some cases, users become more concerned about privacy when directory information becomes more widely or easily accessible. This is the case with Campus Wide Information Systems that get connected to the Internet, allowing people all over the world to locate information that previously was available only to people within a university. These problems also underlie some tension that has surrounded the Netfind user directory service [Schwartz & Tsirigotis 1991]. Originally, "finger" information [Zimmerman 1990] was accessible only within local campus internets. As campuses connect to the Internet, this information becomes accessible to anyone with Internet access. Netfind makes it easy to harness this widely distributed information, and use it as a directory service. Because of this, some people have suggested that Netfind poses a privacy invasion, even though it provides no information that was not already publically available.

What can be done about these privacy problems? A common approach is to enact security barriers, to protect sensitive information. Such barriers help, but often the goals of privacy and security are not well matched. The priority in most security systems is to protect a site's computer systems. Personal privacy is only protected to the extent that it overlaps with this goal. Yet, in most of the above examples, privacy could be threatened without violating a security perimeter. Moreover, in some cases privacy and security are at odds with one another. Every system administrator can tell a story of a time when, in an effort to track down a security problem, they faced a decision about whether to look into a personal mailbox for clues.

Security mechanisms represent policies that have been formalized to the point of explicit controls over what users can do. Often, less formalized policies exist, in the form of proclaimed constraints on acceptable activities. For example, a university computing support group might have written policies describing the conditions under which a system administrator is allowed to inspect a personal mailbox.

Still less formal policies exist concerning what type of behavior is acceptable when accessing machines across the Internet. For example, it is considered acceptable to login and retrieve files from a known remote anonymous FTP file system, but it is not considered acceptable to try to discover anonymous FTP servers by systematically attempting to login to a list of machines. In this case, the policy is really no more than conventional wisdom that has developed over years of collective Internet usage experiences. Sometimes this type of information is written down, in the form of new user guides [Kehoe 1992]. Often, however, it exists simply as folklore.

A difficulty with defining privacy policies is that no governing body has the authority to legislate and enforce global policies. While a number of groups have stepped forward to suggest policies [Cerf 1991, Conklin 1992, Curran & Marine 1992, North American Directory Forum 1992], there are no global privacy standards. Indeed, different countries currently have very different privacy laws and mechanisms [Flaherty 1985]. Even within a country, one can see a range of different directory privacy policy choices. For example, in the United States one extreme is occupied by mailing list brokers, which compile lists of everyone they can locate (assisted by the U.S. Post Office), and offer no assured way for people to modify or restrict access to the information listed about them. Telephone companies have a more moderate policy, listing everyone by default, but allowing individuals to restrict the content or presence of their listings. Still more moderate are the policies put forward by the North American Directory Forum. NADF advocates several principles, including the right to be informed when a person's directory entry is created, the right not to be listed, the right to correct inaccurate information, and the right to remove specific information [North American Directory Forum 1992].

Notice that none of the above policies requires that users give prior consent to be listed - which is often what people upset about privacy intrusions really want. Even the NADF policy allows users to enforce their privacy preferences only after the information has been visible for some initial period of time. Given today's technology, even a short period of time may be enough to spread private information to many derived databases.

There are at least three reasons why privacy policies tend to be weaker than individuals would like. The first is ubiquity of coverage. A user directory is only valuable if it contains listings for many people. Requiring prior approval makes this goal nearly unattainable. The second problem is more vexing: invading privacy can be very profitable (as in the case of direct mail marketing lists). Third, governments often want to maintain more detailed information about people than their citizens would like, in the name of national security.

There are no easy solutions to privacy problems. However, I have three suggestions. First, we need to educate users about the many ways that networked information can invade their privacy. In a sense, networks represent an "electronic society", participation in which brings with it some loss of privacy, just as being a member of any society does.

Second, we need to increase the amount of research and development oriented towards ensuring privacy. Technical conferences on the subject often focus almost entirely on system security, with few papers about privacy.

Third, we need to form policies oriented towards current technology. The world has changed substantially since the U.S. Privacy Act of 1974 was introduced, and that legislation does not effectively address many of the privacy problems that face us today (nor is it effectively enforced [Flaherty 1985]). One possibility would be for the U.S. National Research Council to commission a study parallel to last year's security study [National Research Council 1991], focused on privacy problems raised by networked information.

Or perhaps the Internet Society could commission such a study, being an international organization.

There are many other privacy problems introduced by electronic data processing, beyond issues raised by Internet directory and information services. The interested reader can contact one of the offices of the Computer Professionals for Social Responsibility (the main office being in Palo Alto, California), or see the proceedings of some of the recent conferences on Computers, Freedom & Privacy.

References

[Callimahos 1989]
L. D. Callimahos. Traffic Analysis and the Zendian Problem. Aegean Park Press, Laguna Hills, California, 1989.

[Cerf 1991]
V. G. Cerf, editor. Guidelines for Internet Measurement Activities. Request For Comments 1262, Internet Activities Board, October 1991.

[Conklin 1992]
J. Conklin. Announcement of CREN White Pages Project. Corporation for Research and Educational Networking, Aug. 1992.

[Curran & Marine 1992]
J. Curran and A. Marine. Privacy and Accuracy Issues in Network Information Center Databases. Request For Comments 1355, Aug. 1992.

[Emtage & Deutsch 1992]
A. Emtage and P. Deutsch. Archie - An Electronic Directory Service for the Internet. Proceedings of the USENIX Winter Conference, pp. 93-110, San Francisco, California, January 1992.

[Flaherty 1985]
D. H. Flaherty. Data Protection and Privacy: Comparative Policies. A Report to the Government Information Technology Project, Office of Technology Assessment, U.S. Congress. Technical Proposal US 84-10/9, January 1985.

[Kehoe 1992]
B. Kehoe. Zen and the Art of the Internet: A Beginner's Guide to the Internet. Department of Computer Science, Widener University, March 1992.

[Linn 1987] J. Linn. Privacy Enhancement for Internet Electronic Mail:
Part I: Message Encipherment and Authentication Procedures. Req. For Com. 989, Feb. 1987.

[National Research Council 1991]
National Research Council. Computers at Risk: Safe Computing in the Information Age. National Academy Press, Washington, D.C., 1991.

[North American Directory Forum 1992]
North American Directory Forum. User Bill of Rights for Entries and Listings in the Public Directory. Req. For Com. 1295, North American Directory Forum, Jan. 1992.

[Schwartz & Tsirigotis 1991]
M. F. Schwartz and P. G. Tsirigotis. Experience with a Semantically Cognizant Internet White Pages Directory Tool. J. Internetworking: Research and Experience, 2(1), pp. 23-50, Mar. 1991.

[Schwartz & Wood 1992]
M. F. Schwartz and D. C. M. Wood. Discovering Shared Interests Among People Using Graph Analysis of Global Electronic Mail Traffic. Dept. Comput. Sci., Univ. Colorado, Boulder, CO, Revised October 1992. To appear, Commun. ACM.

[Zimmerman 1990]
D. Zimmerman. The Finger User Information Protocol. Request For Comments 1288, Center for Discrete Mathematics and Theoretical Computer Science, November 1990.