Technological Implications for Privacy

David Kotz
Dartmouth College
dfk@cs.dartmouth.edu

Abstract. The World-Wide Web is increasingly used for commerce and access to personal information stored in databases. Although the Web is "just another medium" for information exchange, the fact that all the information is stored in computers, and all of the activity happens in computers and computer networks, makes it easier (cheaper) than every to track users' activities. By recording and analyzing user's activities in the Web, activities that may seem to be quite private to many users, it is more likely than ever before that a person's privacy may be threatened. In this paper I examine some of the technology in the Web, and how it affects the privacy of Web users. I also briefly summarize some of the efforts to regulate privacy on the Internet.

1. Introduction

As the World-Wide Web is increasingly used for commerce and for access to personal information stored in computerized databases, privacy becomes a significant concern. As noted by Moor [10], computerized information is *greased* so that it flows easily from place to place and person to person, and can be easily correlated and combined with other computerized information to derive still more information. It is clear that the Web eases access to existing data sources, some of which contain personal information (financial records, medical records, criminal records, etc). More subtly, and perhaps more significantly, by moving many human activities (shopping, entertainment, communication, and so forth) onto the computer and computer networks, the Web makes it possible to track an individual's activities more than ever before.

In this paper I examine some of the technological aspects of the Web that may have an impact on privacy, specifically, what Moor calls *informational privacy* [10]. Indeed, the law generally sees the invasion of privacy to be the unauthorized access to or disclosure of information [14]. Why is the World-Wide Web, or the Internet in general, so different from earlier technologies in this respect? "The Web is simply another medium of information distribution and gathering; it differs quantitatively because the volumes are so high and the costs so low" [1]. That is, the low cost of collecting personal data, and the sheer volume of such data on the Web, makes it more likely that individual privacy may be invaded, where before it would have been impractical.

David Kotz   1

For the purposes of this paper, I use Alan F. Westin's definition: "Privacy ...is the claim of individuals... to determine for themselves when, how, and to what extent information about them is communicated to others..." [15]. It is important to consider what "information" individuals may want to control, and to note that the sensitivity of the information varies with the context. Frequently, people distinguish between "personally identifying information," that is, information that can identify a specific individual, and "anonymous information," that is, information that cannot be used to identify them as a specific individual. Normally, people are more sensitive about the former than the latter, but the Web may change that.

It is instructive to consider what it means to *identify* an individual. Most people probably believe that their *name* identifies them, but there are often thousands of people with an identical name. Within the United States, a social-security number is a unique identifier for a specific individual (but not all individuals in the U.S. have a social-security number, and it may be that they re-use numbers after death of an individual). Furthermore, knowledge of a name (say, "Jingle Frobnitz") or even a social-security number (say, 111-222-3333) is, in and of itself, completely useless information. Knowledge of information about an individual is only useful unless I have a use for that information, and the potential to use it. For example, I may be able to use a name to locate a telephone number for that person (or at least, a list of telephone numbers for all people with that name). This additional knowledge is itself useless, unless I have a use for the information, and the potential to use it: thus, if I have a desire to call this person, and a telephone, the information is useful to me. If my goal is to send advertisements through electronic mail, I may have no use for a name or telephone number, but an email address is extremely valuable. If my goal is to cross-sell a customer a shirt after they add a sweater to their virtual shopping cart, knowledge of their gender or age may be more useful than their name or social-security number.

In the end, it appears that the value of information about an individual is directly related to the potential uses of that information. Similarly, the individual's sensitivity to disclosure of certain information about them is directly related to the use (or potential use) of that information, which in turn varies with the context. As such, the line between "personally identifiable" and "anonymous" information is blurry. Furthermore, Web technology makes it increasingly easy to correlate information from multiple sources, so information you provide "anonymously" to one source (such as your age and list of current prescriptions) may be linked with information you provide to another source (such as your email address), allowing someone to send you (or your employer) email about your health-care situation. I discuss the issue of information correlation further in Section 3.

There are three forms of information protection related to privacy on the Web:
- Protection from others eavesdropping in your communications;
- Protection of personal or proprietary information (in databases, caches, or logs);
- Protection from the collection, exchange, and use of personal information, including information about an individual's activity in the Internet.

The first issue is largely solved by the use of encryption. The second issue is largely solved through encryption, and traditional access-control mechanisms (as described in nearly any Operating Systems textbook for computer scientists). This paper focuses primarily on the third issue.

I begin by outlining some of the mechanisms available for accessing or collecting personal information on the Web. Then in Section 3 I examine methods for correlating multiple data sources to provide a more complete view of a person, and in Section 4 I note a few examples of how such data is disclosed to third parties. Finally, Section 5 discusses some of the technical and legal solutions to the privacy issue.

2. Data collection mechanisms

There are essentially three ways to obtain personal information from a World-Wide Web user:

1. **Direct:** record the information they type into forms on web pages (such as a name or address).
2. **Indirect:** record their web-surfing activity, such as the pages they visit, the queries they enter, or items they purchase from a on-line store.
3. **Derived:** correlate data from several sources to infer new facts about them.

First, I describe several mechanisms for data collection.

2.1. Server logs

From the earliest days of the World-Wide Web, browsers have transmitted a few key pieces of data to the web server every time they request a new page or image: time, Internet (IP) address of the client machine, brand and version of the browser software, brand and version of the operating system software, the URL identifying the page requested (of course), and the URL identifying the "referring" web page, that is, the page on which the user just clicked to cause the browser to request this new page.[1] And, from the beginning, web servers have carefully recorded all of this information in a "log" file. Most web-site managers (webmasters) ignore the log, and discard it every few days. A few use software to compute summary statistics, such as the number of requests for each page and the number of requests from each Internet domain.[2]

Note that it is not possible to identify the particular user directly from this information, only the IP address of the client machine. If the client machine is consistently used by only one user, and has been assigned a permanent IP address, then the IP address effectively identifies one individual. Often, however, a single client machine (and single IP address) supports multiple users (not common on Windows or Macintosh machines, but very common on Unix platforms), whose requests are mixed together and not distinguishable by the server. Perhaps the majority of all personal computers today are not assigned a permanent IP address; they are assigned a new address every time they connect to their Internet Service Provider (ISP). As a result, requests coming from one IP address today cannot necessarily be correlated with those coming from that address tomorrow.

Although the server cannot know whether the IP address is shared among several people, some ISPs provide a special call-back service that, given an IP address, returns the

---

[1] Or, in the case of an image file, the page containing the image.
[2] For example, see the statistics for the Dartmouth Computer Science web site: http://www.cs.dartmouth.edu/stats.html.

email address of the individual currently assigned that IP address. This "identd" service can be used by the web server to add email identities to its log.

Although the IP address may not identify the user, individually, it often provides good information about the location of that user. Some advertising agencies offer to target ads based on the top-level domain (.edu, .com, .gov, etc.),[3] while still others offer to target ads to users in a specific geographic region [4], which is of interest to local advertisers.

Finally, FTP (file transfer protocol) services also keep a log of all accesses. While web traffic (using the HTTP protocol) is much more common than FTP traffic an old convention can make the FTP log entries more personally identifiable. FTP requires both a username and a password with every access. Most FTP servers permit access by the "anonymous" user, and ask that you provide your email address as a "password." Any text will suffice, and is recorded. The email-address-as-password convention was formerly considered a courtesy to the administrator of the FTP site, at least, when users had to manually type the word "anonymous" and their choice of a password. Now that most users access FTP through a web browser, which transparently enters "anonymous" and a password, it is best for browsers *not* to transmit the email address unless specifically configured by the user to do so.

So how can a site administrator use the data in the web- and FTP-server logs? Other than the summary statistics mentioned above, the logs can be used for a weak form of tracking individual behavior. That is, by piecing together accesses from the same IP address, and by making the (weakly justified) assumption that all accesses from that IP address represent the same user, you can form a picture of that person's web-surfing activity: which pages are accessed, and in which order, what third-party web sites were used as the "referring" web page, how long did they spend on one page before clicking to the next page, and so forth. That data, once correlated with the identity of the individual, can paint a powerful picture of that individual's habits.

2.2. Forms

Forms are web pages that contain boxes into which the user should type. The user's entries are sent back to a program on the web server, usually when the user clicks a button labeled "Submit" or the like. They can be used to allow the user to type a query for a search engine, to type their name and address as part of a purchase, to type a message to a chat group, and so forth.

A less obvious data-entry form is an "image map". Many "banners" and images on web pages are actually "maps", so that a click on the image will send a request to the web server for another page, with the coordinates of your click sent along with the request. While often used to provide glitzy buttons, image maps are also used to allow you to click on a geographical map or a diagram, leading you to information about that region or about that detail of the diagram.

Needless to say, any personally identifying information (name, address, email address, SSN, credit card number, etc), that is entered into a form, can and likely will be recorded by

---

[3] MatchLogic, among others.

the web server's software.   Less obviously, seemingly "anonymous" data such as the queries entered into a search engine,[4] or the clicks made on an image map, might be recorded and later correlated with other, non-anonymous entries you make.

Form technology can also be used to track the sequence of pages you visit, within a web site. The web server customizes each page that is sent by adding a "hidden" form entry that contains a unique "session number." When you click on the web page, to request a new web page from that site, the session number is sent back to the server along with the request.   (It's not quite this easy, and this use of hidden form data is a bit of a hack. The correct approach is to use *cookies;* see below.) Assuming that the web server generates a new session every time it receives a request with no accompanying session number, and records the session number in its log with all of the accesses, it can more clearly track the activities of any one user (browser) than it could using the IP address alone. This information is part of the web page as it is displayed, however, and disappears when you quit your web browser, so the ability to track a user is limited to a short time.

2.3.   *Cookies*

Cookies are a greatly misunderstood technology, hotly discussed in the past year, and a powerful tool for data collection (among other things). In this section I hope to provide some background, describe how cookies work, and discuss ways in which they can and are being used to collect information about individual Web users.

World-Wide Web "browser" software acts as a *client,* sending requests (in the form of URLs) to Web *servers*, which look up (or compute) the appropriate Web page, and send the page back as a reply. Once the current request is completed (and logged, as I note in the previous section), the Web server forgets everything about this request and goes on to service the next request, most likely from a different client elsewhere.   The server is *stateless,* in that it keeps no state information from request to request. It has no way to know, when a new request comes from a client that made an earlier request, that the request is somehow related to the earlier request.   Many useful applications, from search engines displaying multiple pages of results, to commerce sites providing "shopping basket" in which customers collect items to purchase, require the maintenance of short-term (within the same session) state. Other applications require long-term state to be recorded, such as your preferences for a personalized newspaper, the book reviews you have already seen, and so forth.

Cookies provide a way to supply state information to the Web server. A server that wishes to remember something from request to request can send a few bytes of data back with the web page, in response to your first request. Those bytes of data are called a "cookie", and are recorded by the browser software. The cookie is not a program, and thus cannot execute on your machine. As such, it cannot "do" anything with your machine, despite rumors to the contrary. On your next request to the same Web server, the cookie is sent back along with the request. The server, which is otherwise still stateless, looks at the cookie for information that it can use to help process your new request. On reply, it can change or delete the cookie stored by the browser, or send a new cookie.

---

[4] MatchLogic, an on-line marketer, claims to record and analyze search queries "to help decide which advertisements we present to you...." It appears that they link search-query information to personally identifiable information entered by a user, in response to a sweepstakes entry form, via cookies.

What is in a cookie? A few bytes of data, typically less than 100 bytes, but apparently as much as 4096 bytes. Each cookie has several fields:

- **Name:** the name of the cookie, meaningful only to the server
- **Value:** the value of the cookie, meaningful only to the server.
- **Domain:** the Internet domain (such as `dartmouth.edu`) that created the cookie; this is the only domain to which the cookie will be sent along with future requests.
- **Flag:** true or false: do all machines in the domain get sent this cookie? If the domain is `dartmouth.edu`, and flag is true, then the cookie would be sent along with any requests to `www.dartmouth.edu` or `www.cs.dartmouth.edu`.
- **Path:** the prefix for the set of URLs at that domain for which this cookie will be sent; typically the path is "/", representing all URLs in that domain.
- **Secure:** another flag, typically false, indicating that this cookie should only be sent along secure HTTP (https) connections.
- **Expiration:** the time at which this cookie expires, and can be deleted from the browser's records; the browser is free to discard the cookie earlier if it wishes.

A server may send many cookies, each with a different name. Most browsers will only remember 10-20 cookies from any one server, and at most 300 cookies total. The cookies are recorded in a file on disk, so they can be remembered from one execution of the browser to another.

More technical details about cookies can be found in the cookie specifications (which appear not to be completely followed by current browsers) [10,12] and the Unofficial Cookie FAQ.[5]

As an example of some cookies, here are some of the cookies that I found stored in my browser's records:

Domain	Flag	Path	Secure	Expiration	Name	Value
.nytimes.com	TRUE	/	FALSE	946684801	PW	L=0:/1fl
.nytimes.com	TRUE	/	FALSE	946684801	ID	3/*$:fl
.amazon.com	TRUE	/	FALSE	922089609	group_discount_cookie	F
.amazon.com	TRUE	/	FALSE	2082787207	ubid-main	001-4986362-8617108
.doubleclick.net	TRUE	/	FALSE	1920499140	id	1ee01499

Clearly, these cookies' values have little meaning to the user, but a few things can be gleaned from a casual reading of the cookies.   I have visited both the *New York Times* and Amazon.com Web sites, and each sent me some cookies. Both seem to have sent me a cookie that records some sort of identifier ("ID", "ubid-main"), which is probably used by the server to locate more detailed records about my previous browsing activity in the server's own databases.   Amazon.com records that I do not (F=FALSE) have a group discount.

Every time I return to the *New York Times* or Amazon.com Web site, the appropriate cookies are returned to the server. Thus, it is possible for the server to record (more accurately than through monitoring the IP address, and over longer-term activity than through the use of hidden form fields) the sequence of pages I visit on their server, the time between my accesses, any data I enter into their forms, and so forth. If, at any time, I enter "identifying" (such as my name or email address), all of the recorded information associated with the cookie can now be associated with my name or address. Thus, although it may

---

[5] Unofficial Cookie FAQ: http://www.cookiecentral.com/unofficial_cookie_faq.htm.

appear that some parts of the Web site are possible to access "anonymously," in reality it is possible for the server to relate my activity from one part of the Web site (where I identify myself) to other parts of the site (where I have the impression that I am not identifying myself).

Most notable, however, is the cookie from `doubleclick.net`, a site that I have never intentionally visited. The domain doubleclick.net is an typical of Web-advertising companies. They are an ad agency, paid to create and place ads on web pages by their advertising clients. They pay Web servers to place a link to the ad, which is a URL to doubleclick.net, on the pages produced by that Web server. Thus, whenever I display a page from one of my favorite sites, my browser may pick up an ad from doubleclick.net. In the process, doubleclick.net sends my browser a cookie. The typical purpose of such a cookie is to record which ads have been sent to me before, and which have not, so that they can be sure to send me different ads, or at least, the right sequence or combination of ads, according to the nature of their advertising policy.

2.4.   Other sources of data

There are some other sources of personally identifiable data on the Web. Search engines can search through millions of Web sites for pages that match a particular set of keywords. A search for a person's name will often return links to pages written by or mentioning that person. Since many email-discussion lists are archived to Web pages, it is possible that a search engine can find things you have said on small mailing lists.

One specific search engine is DejaNews, which searches through archives of Usenet newsgroups. There are hundreds of news groups, each on a different topic. Anyone may post to any news group, and any one may read from any news group. The postings are distributed worldwide, copied to all subscribing computer servers. While most servers discard postings after one or two weeks, DejaNews attempts to save a permanent copy of all news postings, to provide a searchable database. After locating one article by a specific person, you can ask DejaNews for an "author profile" of that person, which provides links to all of the articles ever posted by that person. The set of news groups containing posts by a user can provide valuable insight into the person's interests and activity.

2.5.   Existing databases

In addition to the various data that can be collected from your web-surfing activities, there is an increasing amount of personal data, collected from other media, available on the web. These data are provided in searchable databases, often by traditional information agencies, for free or for a fee.

There are several sites that provide telephone-book data, allowing you to search for the address and telephone number (and sometimes email address), given a name. At least one allows "reverse" searches, providing name and address given a telephone number.[6]

More interesting, several sites will provide a wealth of personal information, given a name and social security number. For that, and $139, I was able to ask InfoSpace to run a

---

[6] InfoSpace: http://www.infospace.com/.

"background check" on someone. In 24 hours, they provided me with alternate names, addresses, date of birth, a list of possible relatives, the address and phone number of several neighbors, corporate records, and lists of bankruptcies, real estate owned, watercraft, aircraft, and pilot licenses.

Most of this sort of information comes from public records, which have been scavenged and made easily available through the Internet. Some now question the extent to which public records should be made public, given that the increased accessibility makes abuse more likely.

3. Correlation of personal data

The correlation of data from multiple sources can often provide more information than the sum of its parts, by exposing new patterns or by enabling new uses of previously anonymous data. For example, anonymous tracking data from server logs or cookie-related logging can be tied to an individual who enters personal information into a form. Advertisers, criminals (or criminal investigators), corporate spies, and so forth, may then be able to use the information to target that individual.

It is not difficult to see how this correlation can be accomplished within the context of one Web site. More interesting is the correlation of information across Web sites. Let us return to the `doubleclick.net` cookie seen above. This "third-party" advertising entity may observe, record, and analyze my accesses to many pages from many different Web sites [13]. The *referring URL* information that is sent along with every HTTP request tells them what page I am displaying when I request their ad; if they have ads displayed on many Web sites, they can track my activity across Web sites. If they arrange to share information with the administrators of that Web site, who may have collected my name or other personal information, then they might be able to associate that personal information with my activities on other Web sites, where I had thought that I was completely anonymous. All transparent to me, simply because the sites I visit happen to use the same advertising agency. Indeed, this sort of multi-site analysis is now being arranged by a company called Engage Technologies, for the purposes of "precision targeting" of on-line advertisements.[7]

4. Dissemination of personal data

There are several ways in which Web sites might (and do) use data collected through one or more of the means above. Perhaps the most common use of data is for electronic marketing, typically the placement of "banner" ads on Web pages. Analysis of data scrounged from server logs or from the results of cookie tracking can help classify users into groups that might be appropriate for certain ads. For example, the "referring URL" information could tell a bookseller to suggest certain titles to users arriving from a computer software Web site, and a different set of titles to users arriving from a pornography Web site. Perhaps a certain banner ad attracts clicks when placed on one Web page and not on another. Perhaps users that have read certain combinations of articles in the on-line newspaper would be appropriate targets for certain advertisements.

---

[7] http://www.engage.com/

But more is possible. Some sites entice users to form a "member profile," typically in exchange for free e-mail or other services. This information, in combination with data collected from monitoring users' surfing patterns, is then sold to advertisers and other third parties. Although commercial sites are the most common advertisers, political campaigns are increasingly interested in these targeted advertising avenues [1].

Juno.com asks new members to answer about twenty questions regarding their interests and demographic traits. This information is used to select advertisements that might appeal to you, while you use their Web site (e.g., to read your email). Individually identifiable information is given to third parties only when given permission by the user. Aggregated information is provided to advertisers (e.g., "25 percent of all members who clicked on [your ad] were women with family incomes of more than $50,000"). In their service agreement[8] they retain the right to monitor your usage, including frequency of usage, navigation information, and so forth, and that they may disclose statistical summaries of that information to third parties.

GeoCities.com will disclose aggregated data to other parties, but will only disclose personally identifiable information after receiving specific permission. (Their earlier policy claimed that they would *not* disclose such information without permission, when in fact they were selling such information to others. They recently stopped the practice after action by the Federal Trade Commission [2].)

5. Regulation

Given the potential for abuse of personal privacy by various groups on the Web, it is important to consider mechanisms to regulate the collection and use of personal information. There appear to be three primary approaches: legislative, in the form of governmental regulation; self-regulation, in the form of industry associations and certification programs; and technological, in the form of new technology to mediate the exchange of personal information.

The U.S. government appears to be increasingly concerned about on-line privacy, particularly regarding Web users that are young children (under 12 or 13 years of age). Indeed, the Federal Trade Commission (FTC) recently released a report [6] about a study of 1400 Web sites, only 14% of which provide any sort of privacy notice. A subset of 212 sites, aimed at children, had 89% collecting personal data, but only 54% disclosing that fact; less than 10% of these sites provided for parental control over the data collection. The FTC concluded that self-regulation is not working, so far, and warned the industry to improve its record or to expect the government to step in. They suggest the following four principles:
- Notice: Web sites should provide notice of their information collection and use,
- Choice: users should be able to choose how their information is used, and to whom it is disseminated, and
- Access: users should be able to access the information recorded about them.
- Security: databases of personal information shall be secure against tampering or leaks.
Indeed, a bill was recently introduced in Congress (H.R.2368) to establish voluntary guidelines consistent with the above principles, and (in addition) to restrict e-mail spam.

---

[8] http://www.juno.com/getit1.html, section 4.1, effective date August 3, 1998.

Indeed, Vice President Gore chose to make a significant public statement about on-line privacy on July 31, 1998, one month after the FTC report and the same day that Congress was considering related legislation. He focused on four areas: protection of personal information, particularly medical and banking records; preventing identity theft; protecting children's privacy, on-line; and asking the private sector to protect on-line privacy [8].

Most of Gore's statement reiterated previous policies, or lent support to bills already under consideration in Congress. Many of the on-line privacy initiatives he suggested were aimed at children, prohibiting the collection of data from children under 13 without their parent's consent. Congress has been considering similar legislation in this session, in the *Children's Privacy Protection and Parental Empowerment Act* (S.504 and H.R.1972). Congress has only recently begun to consider adult privacy issues. The*Personal Information Privacy Act of 1997* (S.600 and H.R.1813) primarily deals with uses of the Social Security Number. The *Federal Internet Privacy Act* prohibits federal agencies from making "available through the Internet any record with respect to an individual." (To me, the bill's language appears to make it impossible to name an individual in any web page, which seems overly strong.) The *Consumer Internet Privacy Protection Act*(H.R. 98) addresses the disclosure of "personally identifiable information" by an Internet-service provider to a third party. None of these bills were passed into law, although the Senate has passed a bill banning "identity piracy", in which it becomes a crime to "knowingly … use… someone else's personal identifying information with the intent to commit a federal crime" (*Identity Theft and Assumption Deterrence Act of 1998*, S.512). Some privacy advocates are concerned about some provisions in the *Digital Millennium Copyright Act* (S.2037 and H.R.2281), now law, which implements the WIPO treaty, because it appears to disallow the use of "devices", including software, that circumvent technologies that are used for the protection of intellectual property, e.g., copyright. Thus, it may not be legal to use cookie-blocking software, if some publishers use cookies for copyright enforcement.

Gore encouraged companies that produce digital "profiles" of individuals by correlating multiple data sources to develop adequate self-regulation mechanisms, or be prepared to have the government impose regulations. He supported the Online Privacy Alliance,[9] an industry group that proposes to set privacy regulations, stamp conforming Web sites with their seal of approval, and enforce their regulations by revoking the seal as necessary. TRUSTe , a non-profit initiative, offers a similar stamp of approval on many Web sites.[10] "TRUSTe's position to sites is 'say what you do, then do what you say.'" Thus, rather than imposing particular rules about use and dissemination of personal information, TRUSTe requires its members to disclose their own use and dissemination policies, and then to stick by those policies.

The issue may have practical significance soon, as the European Union is scheduled to implement fairly strong privacy rules in October 1998. The rules would make it difficult or illegal for European companies to do business with U.S. companies that have fewer safeguards on the privacy of employees' or customers' personal information. Although the E.U. has weakened their support for strong encryption, at the request of Britain and France [3], it appears that the U.S. and E.U. fundamentally disagree on the role of government in legislating privacy policy. The U.S. prefers voluntary regulations enforced by a non-profit

---

[9] http://www.privacyalliance.org
[10] http://www.truste.org

industry group, and the E.U. prefers strict governmental regulations [5,7]. Negotiations continue at the time of this writing.

In the midst of all the discussion of governmental regulation, or industrial self-regulation, the World-Wide-Web Consortium (W3C) has formed the P3P Project.[11] *P3P* is the *Platform for Privacy Preferences,* a technology that might be helpful in managing privacy regulations. The users express their privacy preferences to their browser, describing which information they are willing to disclose, and which they are not. When the browser connects to a new site, it asks the site for its privacy policy, which has been encoded by the site's administrators. The browser compares the site's stated policy with the user's preferences, and alerts the user if there is any discrepancy, but does not bother the user if the site's policy is as least as strong as the user's preference for privacy. The goal is to encourage users and sites to codify their preferences and policies, and to automate the users' effort that verify the policy of each site they visit, so that privacy can be maintained without excessive interruption to the users' web surfing.

6. Summary

There are several ways in which an individual's privacy can be threatened while they use the World-Wide Web, but perhaps the most interesting is the collection of data about their activities on the Web, often without their knowledge. Furthermore, the nature of information collection on the Web, and the potential uses (and abuses) of that information, blurs the lines between "anonymous" and "personally identifiable" information, and the contexts in which people may wish to keep certain information private. The Internet community, and world governments, are just beginning to consider regulations (voluntary or otherwise) that might help maintain privacy of the Web's users.

I have collected many relevant links to Web resources on this subject, as well as copies of this paper and the slides used for the conference presentation.[12]

7. References

1. Larry Abramson, "Internet and Politics." *Morning Edition,* National Public Radio, August 17, 1998.
2. Joel Brinkley, "Web Site Agrees to Safeguards in First On-Line Privacy Deal." *The New York Times,* August 14, 1998, page A15.
3. Annamarie Cumiskey, "Connected: Go-ahead for police to tap coded email." *The Daily Telegraph,* September 24, 1998, page 2.
4. "DoubleClick localizes Web Ads", News.com, July 14, 1998.
5. Nancy Dunne, "US-EU in 'productive' talks on internet privacy." *The Financial Times,* July 30, 1998. Edition 2, page 4.
6. Federal Trade Commission, "Privacy Online: a report to Congress", June 1998.
7. Rob Ferguson, "Privacy will be key issue at e-commerce summit." *The Toronto Star,* September 30, 1998. Edition 1, Business section.
8. Al Gore, "Gore Announces New Steps Toward Electronic Bill of Rights." Office of the Vice President press release, U.S. Newswire, Inc. July 31, 1998.

---

[11] http://www.w3.org/P3P/
[12] http://www.cs.dartmouth.edu/~dfk/tangled-web.html

9. Junkbusters, "Junkbusters Submission to the FTC." Section 2.3, document number 32 in project number P954807, Federal Trade Commission. Also available at www.junkbusters.com.

10. D. Kristol and L. Montulli, "HTTP State Management Mechanism." Internet RFC 2109, February 1997.

11. James H. Moor, "Toward a Theory of Privacy in the Information Age." *Computers and Society,* September 1997, pages 27-32.

12. Netscape Corporation, "Persistent Client State: HTTP Cookies. " Seen July 1998 at http://home.netscape.com/newsref/std/cookie_spec.html.

13. U.S. Department of Energy, Computer Incident Advisory Capability, "Internet Cookies." Information Bulletin I-034, March 12, 1998.

14. James Vergari and Virginia Shue, *Fundamentals of Computer–High Technology Law*. American Law Institute, American Bar Association, Philadelphia, PA, 1991.

15. Alan F. Westin, *Privacy and Freedom.* Atheneum, New York, 1967.