IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4429

PERSONAL INFORMATION PRIVACY AND INTERNET TECHNOLOGY

Edward J. Szewczak, Professor
Wehle School of Business, Canisius College, 2001 Main Street, Buffalo, NY 14208
Phone: (716) 888-2239, Fax: (716) 888-2525, szewczak@canisius.edu

ABSTRACT

Personal information privacy is arguably the most important issue facing the growth and prosperity of the Internet, especially of e-commerce. Protecting personal information privacy has ignited a debate that pits privacy advocates against technology growth enthusiasts. This paper explores personal information privacy on the Internet in terms of the technological challenges to personal information privacy facing individuals, businesses, and government regulators.

INTRODUCTION

There is a feeling of online insecurity in the community of Internet users. The results of a 1998 survey conducted by Louis Harris & Associates, Inc. revealed that worries about protecting personal information ranked as the top reason people generally are avoiding the Web (Hammonds, 1998). A 2000 telephone survey conducted by Harris Interactive found that 57% of Internet users favor laws regulating how personal information is collected and used by Internet companies (Green, France, Stepanek & Borrus, 2000). A survey by NFO Interactive (www.nfoi.com) found that the safekeeping of online consumer personal information was the main reason people chose not to shop online. A survey by Jupiter Communications (www.jup.com) found that roughly 64% of respondents do not trust a Web site even if it has posted a privacy policy. The main concern was the handling of credit card data.

On September 9, 1999, *Privacy Times* published the equation "Good Privacy = Good E-Commerce (& Vice Versa)". As events continued to unfold, it became increasingly clear that privacy concerns were plaguing e-commerce. Wall Street began to revalue Internet companies that accumulated customer personal information to target marketing efforts. The FTC told a Senate panel that there were more than 300 online privacy bills to limit the collection and "mining" of personal data pending before state legislatures and Capitol Hill. *Business Week* called the privacy backlash "the privacy penalty" (Stepanek, 2000b). Consumer reaction included an unwillingness to click on Web site banner ads, which in turn lead to advertisers becoming dissatisfied with Web portal effectiveness (Ginsburg, 2000). The director of IBM's Global Trust and E-commerce services unit has been quoted as saying that privacy and security are the largest inhibitors of moving forward for e-business today (Robinson, 2000).

WHAT IS PRIVACY?

In his excellent study on privacy in the information age, Cate (1997) adopted the definition of privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" from Westin (1967, p. 7). Westin/Cate's definition is interesting because it allows for flexibility in discussing privacy within the context of the Internet. Whereas many people worry about divulging personal information electronically, other people seem more than willing to give it away, trading their personal information for personal benefits such as free shipping and coupons (Kuchinskas, 2000). However, personalized service is the main benefit. A Web site can save a shopper time and money by storing and recalling a user's tastes and buying habits (Baig, Spepanek & Gross, 1999). ISPs are willing to allow Web users cheaper access to the Internet provided the users are amenable to having their online behavior tracked for marketing purposes by specialized software. The ISPs share in ad revenues (Angwin, 2000a).

TECHNOLOGICAL CHALLENGES TO PRIVACY

Government regulators and enforcement officials have to consider a host of technological challenges to personal information privacy on the Internet.

Corporate and Government Databases

The practice of gathering personal information about customers and citizens by corporations and governments is well established. Software is available which is dedicated to analyzing data collected by company Web sites, direct-mail operations, customer service, retail stores, and field sales. Web analysis and marketing software enables Web companies to take data about customers stored in large databases and offer these customers merchandise based on past buying behavior, either actual or inferred. It also enables targeted marketing to individuals using e-mail. Governments routinely collect personal information from official records of births, deaths, marriages, divorces, property sales, business licenses, legal proceedings, and driving records. Many of the databases containing this information are going online (Bott, 2000).

Financial information databases. The recent deregulation of the financial services industry has made it possible for banks, insurance companies, and investment companies to begin working together to offer various financial products to consumers. Personal financial information that was kept separate before deregulation can now be aggregated. In fact the ability to mine customer data is one of the driving forces behind the creation of large financial conglomerates. Services can be offered to customers based on their information profiles. Services may also be denied based on information profiles (Consumer Reports, 2000a).

Banks – finance company alliances can disseminate personal information about their customers to third parties without their permission (Consumer Reports, 2000a) and may even decline to alert a customer if someone is snooping in the customer's account (Schoenberger, 2000). Even though companies may choose not to sell personal information to third parties, companies within an alliance may use the data themselves to push financial products

This paper appears in *Managing Information Technology in a Global Economy,* the proceedings of the Information Resources Management Association International Conference.

and services. For example, a financial services company may learn that a person has mutual fund accounts with another company, then call or write the person about their own funds (Sapsford, 2000).

Large credit bureaus such as Equifax and Trans Union have traditionally been a source of information about a person's credit worthiness. Their databases contain information such as a person's age, address, and occupation. Credit bureaus have begun to sell personal information to retailers and other businesses. Equifax has announced its intention to purchase the direct-marketer R.L. Polk & Company, a company that maintains records of consumer's lifestyles and purchase patterns of 105 million households.

Medical information databases. Like personal financial information, medical information is for most people a very private matter. Despite this fact, there is a wealth of personal medical data in government and institutional databases. As Consumer Reports (2000b, p. 23) notes:

> The federal government maintains electronic files of hundreds of millions of Medicare claims. And every state aggregates medical data on its inhabitants, including registries of births, deaths, immunizations, and communicable diseases. But most states go much further. Thirty-seven mandate collection of electronic records of every hospital discharge. Thirty-nine maintain registries of every newly diagnosed case of cancer. Most of these databases are available to any member of the public who asks for them and can operate the database software required to read and manipulate them.

Although many of these government database records are stripped of information which could be used to identify individuals (such as Social Security numbers), it is still possible to link the records to private sector medical records using standard codes for diagnoses and procedures employed by the United States healthcare system. The codes are usually included on insurance claims and hospital discharge records.

The ownership of medical records is not entirely clear. When a record is present in a hospital, the hospital claims ownership. Information delivered to a pharmacy becomes the property of the pharmacy. Even written notes hand written by doctors and nurses are being put into electronic form in the name of faster, more extensive access to needed information.

Much of personal health information which is available to the public is volunteered by individuals themselves, by responding to 800 numbers, coupon offers, rebate offers and Web site registration. Much of the information is included in commercial databases like Behavior-Bank sponsored by Experian, one of the world's largest direct-mail database companies. This information is sold to clients interested in categories of health problems, such as bladder control or high cholesterol. Drug companies are also interested in the commercial databases (Consumer Reports, 2000b).

Medical information databases are at present not available on the Internet, although many are available through private networks. However, this situation is quickly changing. Healtheon and other healthcare companies are competing to get doctors to write prescriptions over the Internet and to persuade people to place their personal health records on the Internet. Healtheon has acquired OnHealth, a consumer health information Web site that has 3.2 million visitors annually, and WebMD and Medcast, which have 100,000 registered physician users, 1.1 million registered consumers, and 2.9 million visitors monthly. It has also acquired medical transaction firms Actamed, Metis, MedE America, Envoy, CareInsite, and Kinetra. It processes an estimated 2,000,000,000 transactions a year, 96% over private data networks and 4% over the Internet (Consumer Reports, 2000b).

E-mail

E-mail accounts for 70% of all network traffic, yet only 10% of it is protected by security measures. Thus it is susceptible to tampering and snooping (Armstrong, 2000b). In many companies, employee e-mail communications are routinely monitored. A 2000 survey of U.S. corporations by the American Management Association found that 54.1% monitored Internet connections, 38.1% monitored e-mail, 30.8% monitored computer files, 11.5% monitored telephone conversations, and 6.8% monitored voice mail (Armstrong, 2000). The survey also revealed that, despite the fact that most companies had policies alerting employees that they were subject to monitoring, 25% surveyed had fired employees based on evidence collected during monitoring (Seglin, 2000). Hackers can also be a problem. Programs can be surreptitiously installed that monitor a user's keystrokes. The keystrokes can be sent across the Internet to a computer that logs everything that is typed for later use (Glass, 2000).

Loss of workday productivity is often cited as the major concern for businesses that monitor e-mail. A single e-mail message heralding a religious holiday was sent to 60,000 employees within Lockheed Martin Corporation, disabling the company's networks for more than six hours (McCarthy, 1999c). Many companies worry about possible litigation stemming from sexually charged e-mail. *Business Week* reports that 70% of employees admit to viewing or sending adult-oriented personal e-mail at work, and 64% sent politically incorrect or offensive personal messages. Xerox, the New York Times, Edward Jones, and First Union Bank have fired employees who sent sexually offensive messages via e-mail. Chevron Corporation and Microsoft Corporation have settled sexual harassment lawsuits as a result of internal e-mail that could have created hostile work environments (Conlin, 2000). Companies are also concerned with activity which may expose the company to breach of contract, trade secret, and defamation lawsuits (Armstrong, 2000).

Employee's invasion of privacy claims have not been upheld in the United States courts which argue that, since employers own the computer equipment, they can do whatever they want with it. Interestingly an attempt has been made recently to challenge the tendency of the courts to side with the employer. Using the National Labor Relations Act, a depression-era law that protects the rights of workers to communicate freely with one another about work terms and conditions, lawyers representing a Timekeeping Systems, Inc. programmer reversed his firing on the basis of an e-mail message questioning the company's new vacation plan (McCarthy, 2000).

In an ironic turn of events, Burlington Northern Santa Fe Corporation sent company e-mail to the screen of a former employee, who had been fired for subordination some years earlier. The e-mail message contained an attached file containing the names, salaries and Social Security numbers of roughly 800 railroad employees. The former employee planned to publish the data on the Internet, so the company sued citing Minnesota tort law under which individuals can be found liable for "publication of private facts." The judge decided in favor of Burlington Northern Santa Fe Corporation. However the Eighth U.S. Circuit Court of Appeals set aside the ruling (Orey, 2000).

Wireless Communications

A monitoring operation run by the U.S. National Security Agency called Echelon uses satellite technology to listen in on virtually all international and (to a limited degree) local wireless communications, including phone calls, faxes, telexes, e-mail and all radio signals including short-wave, airline, and maritime fre-

quencies. The operation listens for certain target words. When a target word is encountered, the transmission is sent to humans for analysis. Echelon is designed primarily for non-military targets, including governments, organizations and businesses around the globe (Port & Resch, 1999). It even has a specially adapted submarine that taps into cables on the ocean floor (Dornan, 2000).

Wireless advertising promises to pose a host of challenges for privacy advocates. Wireless service providers know customers' names, cell phone numbers, home and/or office addresses, and the location from where a customer is calling as well as the number a customer is calling. Each phone has a unique identifier that can be used to record where in the physical world someone travels while using the cell phone (Petersen, 2000). In 1999 Sprint admitted that its "Wireless Web" technology was revealing customer phone numbers to every site they visited (Dornan, 2000). In addition, beginning in October 2000, the Federal Communications Commission will require cell phone service providers to be able to identify the location of a caller who dials 911, the emergency number. Most likely cell phone manufacturers will meet this requirement by embedding a Global Positioning System (GPS) chip in all cell phones. Since a cell phone service provider can track the location of a 911 call, it will also be able to track the location of any other call as well. Companies such as Profilium Inc. are being created to target advertising at people based on their location. Since people usually have their cell phones available most of the time, marketers see an opportunity to reach potential customers, provided the wireless service providers are willing to share their customer information with them (Petersen, 2000).

Clickstream Tracking

As with e-mail technology, productivity and legal liability concerns are also paramount in companies' decisions to track the behavior of employees when using the Internet. *Business Week* reports that 57% of employees say that Web surfing decreases their productivity, 37% surf the Web at work for personal reasons "constantly," and 29% have been caught at work surfing non-work related sites. The most commonly visited sites at work feature pornography, though other activities also occupy workers' attention. Up to 70% of Charles Schwab & Co. customers do online trading from their office desks. Most Hallmark.com transactions occur during traditional work hours (Conlin, 2000).

Tracking employee behavior on the Internet is becoming common practice. Software programs have been specifically designed to monitor when employees use the Internet and which sites they visit. Telemate.Net can examine company network activity and produce reports identifying and ranking the company's heaviest individual Internet users. It lists the sites most visited by members of the whole company or by members of individual departments within the company, and if desired can list sites visited by individual employees and rank them by roughly two dozen categories. Logs can reveal who went to which sites at what times (McCarthy, 1999b). Using internal programs with names like Merlin and Qwizard, Lockheed Martin Corporation tracks employee usage of step-by-step training sites on ethics and legal compliance. The monitoring programs alert managers about employees who haven't taken the required sessions (McCarthy, 1999a).

Internet companies monitor Internet user behavior by a number of means, primarily to gather data about shopping and buying preferences with a view toward developing "user profiles." These technological means include capturing and examining environment variables, cache memory, and cookies (http://www.cnil.fr/traces).

Environment variables contain data about a user's system configuration and site last visited. These variables include a user's domain name, system address, IP (Internet Protocol) address, operating system version, browser version, and URL of the last site visited. This data is transmitted with each packet of data transmitted to an Internet server, where it is extracted by a CGI (Common Gateway Interface) script (program). This capturing of data happens without the user's explicit consent. This data may be saved by the server in a file along with data about any file that the server may have sent to the user in response to the user's request.

Cache memory is a commonly available technology that was developed to enhance file download time and maximize network performance. When a user requests access to a Web site, the browser checks a directory on the user's hard drive to see if the Web page has been loaded previously. If not, it carries out the request but when the page arrives, it records it on the user's hard disk and displays it on the screen. The next time the same request is made, the browser reads the page from the hard disk and saves network resources by not requiring an additional server transmission. The cache memory is not hidden and may be accessed by an outside observer using another computer, thereby revealing the sites that have been visited.

Cookies are text files created by a Web server and stored on a user's hard disk. A cookie is a set of fields that a user's computer and a server exchange during a transaction. The server may change or suppress the contents of a cookie it has created. Web servers work with ad placement companies that resell advertising space from popular sites. These companies maintain large databases in which are recorded details about who looks at which pages. When a user connects to a Web site, the browser checks the cookies on the hard drive. If a cookie matches the site's URL, the browser uploads the cookie to the Web site. With the information contained in the cookie, the site can run programs which personalize site offerings and/or track the user's activity while online.

Cookies have received a great deal of attention from privacy advocates. Among the concerns that have been raised are the following (http://www.anu.edu.au/people/Roger.Clarke/II/Cookies.html):

- Cookies are a surreptitious feature, and were introduced surreptitiously;
- Cookies take advantage of the user's facilities without express or even implied permission to do so;
- The Web server causes a cookie to be read without reference to any setting in the user's Web browser – Read access is ON with no opt-out mechanism;
- The Web browser settings default to writing all cookies, without asking the user what settings are desired.

The Electronic Privacy Information Center (EPIC) filed a complaint with the FTC concerning the information collection practices of DoubleClick, Inc. and its business partners. The complaint alleged that DoubleClick unlawfully tracked the online activities of Internet users through the placement of cookies. EPIC also alleged that DoubleClick combined users' Web surfing records with detailed personal profiles contained in a national database following DoubleClick's merger with Abacus Direct, the country's largest catalog database firm.

DoublesClick's use of banner ads to track online activities is the most commonly used method. But Consumer Reports (2000a) indicated a clandestine approach to tracking which was implemented at Web portals Lycos and Excite on behalf of DoubleClick and Matchlogic. Cookie-like Web bugs are embedded as a transparent GIF (Graphics Interchange Format) the size of single pixel on the user's screen. Since most Web browsers do not alert users to the placement of a cookie by default, most users would be unaware that servers had planted a cookie and were tracking their move-

ments on the Web.

Hardware and Software Watermarks

Hardware and software identifiers ("watermarks") can also be used to identify individual users.

In January 2000 Intel Corporation announced it would include a unique Processor Serial Number (PSN) in its new Pentium III microprocessor chips. The rationale for the PSN was that it was to be used for authentication purposes in e-commerce insofar as the PSN would be linked to a person's real-world identity.

Privacy advocates countered that the PSN could be read remotely by Web sites and other programs and used to link users' activities on the Internet for marketing and other purposes such as identifying users seeking access to chat rooms. As the Big Brother Inside Homepage explains (http://www.bigbrotherinside.org) :

> The PSN would likely be collected by many sites, indexed and accumulated in databases. Unlike cookies, which are usually different for each web site, the PSN will remain the same and cannot be deleted or easily changed. The advertising and marketing industries have been strongly advancing technical means of synchronizing cookies so that information about individual consumer behavior in cyberspace can be shared between companies. We believe that a hardware PSN used in the majority of computers would quickly be put to this purpose. The records of many different companies could be merged without the user's knowledge or consent to provide an intrusive profile of activity on the computer.

Even as a security device the PSN was inadequate. Hackers could easily forge a PSN, thereby negating its authentication rationale. Software patches to disable the PSN have proved inadequate. In April 2000 Intel Corporation decided not to include the PSN in its forthcoming 1.5 GHz Williamette chip. However, in addition to the Pentium III machines with the PSN, there are some Pentium II and Celeron processors that have the PSN.

Every Ethernet card used in computer communications has its own MAC (Medium Access Control) address, a 48-bit number sent in the header of every message frame. As the Ethernet standard evolves into a wide-area communications protocol, this identifier may become of increasing concern to Internet users intent on protecting their privacy (Dornan, 2000).

Microsoft Corporation includes a unique numeric identifier into every copy of its Office program. When a Microsoft Office document is created, it is watermarked with this unique identifier. The creator of the Melissa virus was apprehended when he posted documents to a Web site frequented by virus makers. Authorities used the watermark found in the Melissa virus to match the watermark found in the documents (http://www.forbes.com/Forbes/99/1129/6413182s1.htm).

Biometric Devices

Various devices are available that identify people through scans of their faces, hands, fingers, eyes, or voice recognition. Biometric devices create a statistical profile by assessing a number of biological characteristics. As the equipment used to take the measurements decreases in cost, it becomes economical to scan millions of faces and other characteristics into a computer database. Digital photography adds to the growing volume of non-text data about people. Privacy advocates object to the fact that much of the measurement taking happens without the knowledge or explicit cooperation of a subject, which can lead to abuses of the technology. A spokesperson for the Electronic Frontier Foundation has noted that a bank that has collected face scans of ATM customers could sell this information to another company for a purpose not related to banking (Stepanek, 2000a). Though not as simple as text data, biometric data can be transmitted on the Internet with little difficulty.

CONCLUSION

The issue of personal information privacy and the Internet continues to be debated within the community of Internet users. The concerns of privacy advocates conflict with the concerns of technology growth advocates. Clearly the challenges to personal information privacy posed by the various forms of Internet technology are not the result of the technology itself. Rather it is the uses of the technology that poses the threat to the integrity of personal information privacy. Assuming the privacy concerns of individuals do not bring e-commerce to its knees, Internet technology will continue to grow and be used to bolster the growth of e-commerce. Any laws that are passed must take into account the evolving nature of technology, while at the same time respect the personal information privacy values of individuals. As members of the Internet community, we can only hope that the future of e-commerce will not be embroiled in litigation, which will bring perhaps debilitating expenses to both e-business and individuals alike.

REFERENCES

Available from the author.