CHAPTER 11

Privacy

When the Web started, one of the things holding it back was often people's unwillingness to be open about their workings—their sources and reasons behind their work. I found this frustrating myself, and would carry the banner for openness of information while I was promoting the Web as one way of fostering this openness. However, I rapidly separated the two, as the Web does not and should not imply that all information must always be shared. To maintain integrity, a group needs a respected border, which in the Web is a border of information flow. Groups need to be able to talk among themselves, and have their own data when necessary.

Perhaps the greatest privacy concern for consumers is that, after they have ordered enough products, companies will have accumulated enough personal information to harm or take advantage of them. With consequences ranging from the threat of junk mail to the denial of health insurance, the problem is serious, and

two aspects of the Web make the worry worse. One is that information can be collected much more easily, and the other is that it can be used very easily to tailor what a person experiences.

To see just what can happen to my personal information, I have traced how some online purveyors have used my address. When I provide my address to a Web site, I put a bogus line in it, like an apartment number. Their computer regurgitates it verbatim, so I can tell, when I get junk mail later, who has furnished my address.

There are more threatening scenarios. Burglars could find it very handy to know who has been buying what recently. More likely is the sort of abuse that occurs when a doctor divulges someone's medical condition to the patient's insurance company to justify the claim. Two years later, the insurance company picks the information out of its database when a prospective employer wants to check that person's record. The person doesn't get the job because of a previous medical condition and never even knows what happened.

Software can even track the pattern of clicks a person makes on a Web site. If a user opens an online magazine, the publishers can watch which items he reads, tell which pictures he calls up and in what order, and extract information about him that he would never volunteer on a form. This is known as "click stream" information. Net Perceptions, started by a former head of Microsoft's programming languages division, is one firm that makes software that companies can use to monitor all sorts of online behavior, from the amount of time a visitor spends reading about a product to what pages they print on their printer.

If an advertiser runs ads on different sites and finds a person's click stream on a certain selection of the sites, it can build up an accurate profile of sites that person visits. This information can then be sold to direct marketers, or whomever. A famous cartoon drawn early in the Internet's life depicts two dogs sitting at a computer. One explains to the other, "The great thing about the

Internet is no one knows you're a dog." It has been followed recently by another cartoon in which one dog has clicked to a page with a picture of dog food. Because of this, the server now does know it's a dog. Pretty soon the server also knows it's a dog that prefers a certain brand of dog food, elm trees, and Siamese cats.

In the basic Web design, every time someone clicks on a link, their browser goes from server to server afresh, with no reference to any previous transactions. The controversial tool for consumer tracking that changes all that is the *cookie*. A cookie is just a code such as a reference number or account number that the server assigns to the browser so as to recognize it when the same person returns. It is much like getting an account number when opening a bank account. The cookie is automatically stored on the consumer's hard drive, with or without his knowledge, depending on his preferences.

Most transactions between a consumer and a store involve some continuity, and the cookie makes it possible to accumulate things in a shopping cart, or send items to the same address as last time. Normally, merchants we trade with know what we have bought, and with whom we bank, and where we live, and we trust them. The fact that cookies are often installed on a person's hard drive, and talk back to the server, without any form of permission is also valuable: It's the difference between going into a store and being recognized as creditworthy, and going in and having to fill out identification forms all over again.

However, some commentators see cookies as entirely evil. By default, most browsers accept all cookies automatically, but then again most also offer the user the option of prompting them with an alert notice before the computer accepts a cookie, or of simply refusing it. The problem is not in the cookie itself, over which the user has control. The problem is that there is no knowing what information the server will collect, and how it will use that information. Without that information the user can make choices

based only on fear and doubt: not a stable basis for building society on the Web.

A Web site also can change chameleon-like according to who is looking at it, as if it were a brochure being printed for that one person. Imagine an individual visiting the Web page of a political candidate, or a controversial company. With a quick check of that person's record, the politician or company can serve up just the right mix of propaganda that will warm that particular person's heart—and tactfully suppress points he or she might object to. Is this just effective targeted marketing, or deception? It depends on whether we know it is happening.

Europe has tried to solve part of this problem with strict regulation. European companies have to keep secure the information they hold on customers, and are barred from combining databases in ways that are currently quite legal in the United States. Consumers in Europe also have the right to look at and correct databases that contain information about them. In the United States, laws that protect consumers from having their information resold or given away are very weak. The government has hoped that some sort of self-regulation will come into force.

The good news is that the Web can help. I believe that the privacy I require of information I give away is something I ought to have choice about. People should be able to surf the Web anonymously, or as a well-defined entity, and should be able to control the difference between the two. I would like to be able to decide who I will allow to use my personal information and for what.

Currently, a responsible Web site will have a privacy policy one click from the bottom of the home page. One site might sell any information it gets to direct-mail firms or advertisers. Another may record every page a visitor views. Another might not distribute any information under any circumstances. I could read this carefully and decide whether to proceed, but in practice I usually don't have time to read it before rushing in.

The next step is to make it possible for my browser to do this for me—not just to check, but to negotiate for a different privacy policy, one that will be the basis for any subsequent release of information. With privacy software, a Web site provider and browser can do just that.

Consider a company selling clothing over the Internet. It might declare its privacy policy as follows: "We collect your name, age, and gender to customize our catalogue pages for the type of clothing you are likely to be interested in and for our own product development. We do not provide this information to anyone outside our organization. We also collect your shipping information. We may distribute this information to others."

For these things to be negotiated automatically, the preferences set by a user and the privacy policy have to be set up in machine-readable form using some common set of categories for different sorts of data and different ways of using it.

The World Wide Web Consortium is creating a technology that will allow automatic negotiation between a user's browser and a store's server, leading to an agreement about privacy. The Platform for Privacy Preferences Project (P3P) will give a computer a way of describing its owner's privacy preferences and demands, and give servers a way of describing their privacy policies, all done so that the machines can understand each other and negotiate any differences without a person at either end getting involved.

I believe that when a site has no privacy policy there ought to be a legally enforced default privacy policy that is very protective of the individual. Perhaps this view shows my European roots. And it may sound counter to my normal minimalist tendencies. But lack of such enforcement allows a company to make whatever use it can of whatever private data it can somehow extract.

In 1998 the Federal Trade Commission did a survey of Web sites and found that very few had a privacy policy, including sites that took information from children. The findings were so dramatic

that President Clinton called a two-day Internet privacy meeting in Washington with industry and government officials. The results also prompted the Federal Trade Commission to consider regulating privacy policies.

As is so often the case, the possibility of regulation has prompted industry to make some moves toward self-regulation. In June 1998, Christine Varney, a former FTC commissioner, put together a group of about fifty companies and trade groups called the Online Privacy Alliance. Members included AOL, AT&T, Microsoft, Netscape, the Direct Marketing Association, and the U.S. Chamber of Commerce. They said they would clearly reveal what information they collected on all their various Web sites and how it would be used. They also said they would give consumers some choice about how personal data could be used, including the ability to not allow their information to be sold to third parties. The Better Business Bureau Online is also addressing the matter with an endorsement service—a privacy seal it will grant to worthy Web sites. The program features privacy-standard setting, verification, monitoring, and review of complaints.

Some regulators maintain that since there is no mechanism for enforcement, this kind of effort does not go far enough. Tighter control, they say, is needed, especially when it comes to protecting information about children. They maintain that any abuse of information about adults or children should be illegal. But the Online Privacy Alliance is a good start, at least in creating a system of endorsements, which will cause more consumers to gravitate toward sites that comply. This will put pressure on others to do the same. Ideally, such groups will set privacy practices that will be automatically checkable with P3P.

Of course, any privacy negotiation is only as trustworthy as the site's proprietor. However, if a company has, through its Web server, made an undertaking to preserve privacy, and broken that undertaking, then it has acted fraudulently. There are conventional laws to deal with this transgression. Software can't solve

this problem. And it should not be up to the consortium or any other technical body to solve it.

Perhaps the most notorious violation of privacy over the Web was the sudden release late in 1998 of details from the U.S. Independent Council's report about President Clinton's sexual activities. This information was purposely exposed to millions of people, contrary to many people's concepts of respect for the individual or family. We can use the power of the Web to connect anything and everything to great effect, or to do devastating damage. Episodes like this help us recognize how rapidly the widespread distribution of information could cripple our society—and each of us personally—if absolutely all information remains public.

No one will take part in the new weblike way of working if they do not feel certain that private information will stay private. In a group, they will also remain on the sidelines if they feel that what they say or write will not remain confidential, or if they can't be sure of whom they are communicating with.

Public key cryptography (PKC) offers one way to achieve the four basic aspects of security: authenticity, confidentiality, integrity of messages, and nonrepudiatability. Each person has a number that everyone knows (the *public key*), and another, related number that no one else ever has (the *private key*). Devised more than two decades ago, PKC provides a form of encryption in which an outgoing message is scrambled according to the receiver's public key. The scrambled message can then be decoded only by a receiver who has the unique matching private key to unlock it. A leading form of public key cryptography is RSA, named after its developers, Ron Rivest, Adi Shamir, and Leonard Adleman, all of whom were at MIT's Laboratory for Computer Science in 1977 when they invented it.

Deducing whether someone or someplace is authentic begins with common sense. If a Web site offers a deal that seems too good to be true, it probably is. Tougher, however, is figuring out

whether the Web site of a well-known clothing store is indeed operated by that store. Anyone can make a site that looks like a clothing store. Crooks could even have an elaborate impostor site that takes an order, passes it to the real store, sends the store's communication back, and in the meantime steals the credit-card number. And unlike a physical facade, the fake store will look and feel indistinguishable from the real one. Currently there is some attempt to make the domain name system more secure, but at the moment authenticity relies mainly on the security from intrusion of the domain servers (which tell the browser where, for example, *www.acme.com* is on the Internet) and the connections between them. Public key authentication would be much better.

Confidentiality consists in knowing that no one else can access the contents of a communication. Once again, criminals or spies can intercept a communication to a clothing store and skim off credit-card numbers being sent electronically, or eavesdrop on supposedly private conversations between people in a group. Encryption technology prevents this by scrambling the messages. Anyone browsing a site whose URI starts with *https:* is using an encryption technology called Secure Socket Layer. Normally, however, cryptography is used only to make sure no one except the server can read the communication—not to verify that the server is really who it says it is.

The integrity of messages involves making sure no one can alter a message on the Internet without being detected, and non-repudiatability means that if I have sent a message, I can't later maintain that I did not. PKC provides technology to assure these, too. If I use the software to add to a message I send (or a Web page I write), a number at the end called a *digital signature* allows the receiver to verify that it was I who sent it and that it has not been tampered with. The consortium has a project for applying digital signatures to documents.

If PKC is so well understood, why are we not using it? One reason is the government's fear of loss of control. It is easy to use,

and virtually impossible to crack—so impossible, in fact, that since its development more than twenty years ago the U.S. government has blocked the export of strong cryptography by classifying it as "munition." Some other governments have reacted in similar ways, blocking export, or banning its use, for fear that terrorist groups will be able to communicate without government being able to tap into their conversations.

The counterargument points to George Orwell's vision in his book *1984*, in which the National Security Agency becomes Big Brother, able to monitor a person's every move. It argues that without the basic right of the citizen to discuss what he or she wants, the people are left at the mercy of potential dictatorial tendencies in government.

The balance in governmental power is always a tricky thing. But the debate is almost moot in this case, because encryption technology has been written in many countries of the free world. The U.S. export ban frustrates people who simply want, say, to buy clothes from another country. It infuriates software manufactures who have to make two versions of each product, one with strong PKC and the other with a specifically weakened version for export, and then devise ways of trying to prevent the strong one from crossing borders. It hobbles the Open Source community, in which distribution of the source code (original written form) of programs is a basic tenet. To ridicule the export law, PKC programs have been printed on T-shirts, and in machine-readable fonts in books—which cannot be subject to export controls.

There is another reason why PKC has not been adopted: It can only be used in conjunction with a system for telling your computer which public keys to trust for which sorts of things. This is, of course, a very important but also difficult thing to express. An individual's ability to express trust is essential, because without that trust many uses of the Web, from collaborative work to electronic commerce, will be socially impossible.

Authenticity and confidentiality are not problems new to the Web. They have been solved, in principle, for electronic mail. Pretty Good Privacy (PGP) and Secure MIME are two standards for digitally signing mail (to authenticate the person who sent it) and encrypting it (to stop anyone else from reading it).

PGP is more or less a grassroots system. It is a *web of trust*. An alternative, Public Key Infrastructure (PKI), is basically a tree-like way of doing things. In either PGP or PKI, a user's computer associates a key with a person by holding a file called a certificate. It typically carries the person's name, coordinates, and public key. The certificate itself is digitally signed with another public key of someone else the user trusts. He knows that it is the other person's key because it says so on a certificate that was signed with a key of a different party he trusts. And so on, in a chain.

The social structure assumed by PGP is that chains of trust will be made through anyone—a person's family, friends, college, employer. If an individual was authenticating a message from a colleague, he probably would use a certificate signed by their common employer. That is the path of trust.

The PKI system, planned by industry to enable electronic commerce, assumes that people trust just a few basic "roots" from which all authority flows. A few certificate authorities delegate the right to issue certificates to their commercial partners. They in turn can delegate the right to issue certificates to other, smaller authorities. There is a tree, and money and authority flow up and down it.

Browsers are now slowly being equipped to work with the Public Key Infrastructure. If I open the browser preferences on my Internet Explorer now, I see that I can chose to accept certificates signed by Microsoft, ATT, GTE, MCI, Keywitness Canada Inc., Thawte, and Verisign. In the equivalent list in Netscape, I see ATT, BBN, BelSign, Canada Post, Certisign, GTE, GTIS, IBM, Integrion, Keywitness, MCI Mall, Thawte, Uptime, and Verisign.

All these certificate authorities will vouch for the identity of people and their keys. They generally sell certificates, which expire after a certain number of months. But I don't see a button to set myself up to issue a certificate to a friend or relative whom I also trust. PGP would allow this.

The Web worked only because the ability of anyone to make a link allowed it to represent information and relationships however they existed in real life. The reason cryptography is not in constant use in representing trust on the Web is that there is not, yet, a weblike, decentralized infrastructure.

The PGP system relied on electronic mail, and assumed that everyone held copies of certificates on their hard disks. There were no hypertext links that allowed someone to point to a certificate on the Web. Clearly, it should be much easier to introduce a Web of Trust given the Web.

I mentioned that both PGP and PKI made two assumptions: that we trust a person, and that if we do, we just have to link a person with a key. Many pointless arguments and stalling points have involved exactly what constitutes a person, and how to establish the identity of a person. In fact, in most situations it does not matter who the person "is" in any unique and fundamental way. An individual is just interested in the role the person plays, which is represented by a public key. All we need to do is find a language for talking about what can be done with different keys, and we will have a technical infrastructure for a Web of Trust. If we play our cards right, the work at the consortium in languages for the Web (which I will describe in chapter 13) will end up producing a Web of Trust. Then the Web and the Web of Trust will be the same: a web of documents, some digitally signed, and linked, and completely decentralized. The consortium will not seek a central or controlling role in the Web of Trust; it will just help the community create a common language for expressing trust.

The Web of Trust is an essential model for how we really work as people. Each of us builds our own web of trust as we mature from infancy. As we decide what we are going to link to, read, or buy on the Web, an element of our decision is how much we trust the information we're viewing. Can we trust its publisher's name, privacy practices, political motivations? Sometimes we learn what not to trust the hard way, but more often we inherit trust from someone else—a friend or teacher or family member—or from published recommendations or endorsements by third parties such as our bank or doctor. The result of all this activity creates a web of trust in our slice of society.

Automated systems will arise so negotiations and transactions can be based on our stated criteria for trust. Once we have these tools, we will be able to ask the computer not just for information, but why we should believe it. Imagine an Oh Yeah? button on a browser. There I am, looking at a fantastic deal that can be mine just for the entry of a credit-card number and the click of a button. I press the Oh Yeah? button. My browser challenges the server to provide some credentials. Perhaps this is a list of documents with digitally signed endorsements from, say, the company's bank and supplier, with the keys to verify them. My browser rummages through these with the server, looking to be convinced that the deal is trustworthy. If it's satisfied, good for me, I got a deal. If not, I probably just saved myself some grief.

It would be wrong to assume that the Web of Trust is important primarily for electronic commerce, as if security mattered only where money is concerned. The Web is needed to support all sorts of relationships, on all levels, from the personal, through groups of all sizes, to the global population. When we are working in a group, we share things we would not share outside that group, like half-baked ideas and sensitive information. We do so because we trust the people in the group, and trust that they won't divulge this information to others. To date, it has been difficult to manage such groups on the Web because it is hard to control access to information. The Web of Trust has to evolve before the Web can serve as a true collaborative medium. It has to be there before we can trust automated agents to help us with our work. These developments, which I discuss in the next two chapters, are for me the next most important developments for the Web as a whole.