Technological tools to protect online privacy perform many different functions. A variety of tools are available for encrypting files and email, establishing secure channels to web sites, and establishing encrypted "tunnels" between two computers on the Internet. These tools prevent eavesdropping and protect data from unauthorized access. In addition, anonymity tools are available that prevent online communications from being linked back to a specific individual and prevent eavesdroppers from learning with whom an individual is communicating. Some tools allow users to build anonymous, yet persistent, relationships with web sites, thus allowing sites to track user behavior or provide customized services without building identifiable user profiles.

As popular web browsers have added features that make it easier for web sites to track an individual's browsing behavior, tools have been developed to disable these features. For example, a variety of tools are available that can block or give users more control over the use of web "cookies" — bits of information that a web site can store on a user's computer that will be transmitted automatically back to that site every time the user returns. In addition, many anonymity and cookie blocking tools also block the automatic transmittal of the HTTP_REFERER header, which tells web sites the address of the last site the user visited.

The Platform for Privacy Preferences will enable a new kind of privacy tool that will assist Internet users in finding out about web site privacy practices and making decisions based on these practices. A number of identity management tools and services — sometimes referred to as infomediaries — also claim to assist users in similar ways.

In this paper I provide a brief overview of the various technological tools that can help people protect their privacy online.

**Anonymity and Pseudonymity Tools**

Internet users often engage in interactions with web sites that do not require or even benefit from any exchange of personal information. Users who simply want to browse web pages may have no need to reveal personal information. However, by default, information is revealed automatically by their browsers.

A number of tools have been developed to help Internet users surf the Web
anonymously. These anonymizing agents focus on ensuring that requests to Web sites cannot be linked to an IP address from which a user can be identified.

**Anonymizing Proxies**

Anonymizing proxies are services that submit requests to web sites on behalf of users. Because all requests are submitted by the proxy, the only IP address revealed to web sites is that of the proxy. However users of these services are not anonymous to the proxy, nor to their own Internet service providers, who may log their users’ Web activities. Before using an anonymizing proxy service, users should read the services’ policies to find out what information the service logs and to whom they may disclose that information. Users should also realize that the service may be compelled to share their log files with law enforcement officers, even if there is no mention of this in a service’s policies. And, of course, an unscrupulous service may not actually follow its policies at all!

In addition, users must disable the use of Java and JavaScript in order to ensure their anonymity when using most anonymizing proxies. Some anonymizing proxies claim to automatically block potentially hostile Java and JavaScript programs, but most require users to disable them in their browsers.

A wide variety of subscription and free anonymous proxy services are currently available. One of the best-known anonymizing proxy services is the Anonymizer [1]. Privada [2] is a commercial anonymizing proxy that offers services to ISPs. Anonymity 4 Proxy [3] is software that includes a database of hundreds of public anonymous proxies. It allows users to select a single proxy or use different proxies for each request.

**Mix Networks and Similar Web Anonymity Tools**

Several anonymity tools have been developed around the concept of mix networks [4]. A mix network is a collection of routers called mixes that use a layered encryption technique to encode the path communications should take through the network. In addition, mix networks use other techniques such as buffering and message reordering to further obscure the correlation between messages entering and exiting the network.

In order to better understand how mix networks work, let us imagine a hypothetical secure parcel delivery service called SPS. SPS customers can send parcels to anyone in the world without revealing their identity to the recipient, SPS, or any other party. SPS has a network of branch offices around the world. Parcels are sent from office to office, until finally they are sent in the postal mail to their destination. To send a secure parcel, a customer goes to a local SPS office and selects a set of branch offices through which to route her package. To maximize security, she can spin a dial and randomly select offices; this will of course usually result in an indirect route for the package, but SPS never promises a prompt delivery! Once the customer selects her route she picks up a set of nesting metal boxes and a special lock for each office she has selected for her route. The locks slip onto the boxes like the combination locks used on gym lockers. Once locked, they can only be unlocked by employees of the designated office. The customer marks the package with the address of the recipient. Then she places it inside
the smallest metal box and locks it with the lock of the last office on the route she selected. She marks the outside of the box with the name of that office. She then places the locked box inside a slightly bigger metal box and locks it with the lock of the next-to-last office on the route she selected. And she marks the outside of the box with the name of that office. She continues this process until she has used up all the locks and boxes. She pays the clerk and deposits her locked and boxed parcel in a bin for delivery.

When a truck arrives full of parcels, the clerks unload the truck and unlock the boxes around each parcel. If the parcel contains yet another locked box, the clerk puts it in a delivery bin, mixing it in with the parcels brought in to that branch office by customers. Otherwise, the clerk sends it to the post office for delivery to the final recipient. The truck is then reloaded with waiting parcels from the delivery bins until there is no more room. If there are not enough parcels to fill the truck, the clerks put locks on empty boxes and place them on the truck. If there are too many parcels, some will have to wait for the next truck to arrive. Because the new parcels brought in are mixed with the parcels already waiting, parcels are not necessarily shipped out in the order in which they arrived.

Clearly this is not a very efficient way to run a shipping business, but it does achieve the desired security results.

- By allowing customers to select random routes and place their parcels in locked nested boxes corresponding to these routes, SPS ensures that when a parcel arrives at a branch office, that office knows only the branch it came from and the branch to which it is headed. It has no way of knowing the identity of the sender or the recipient. (The clerks at the first office on the route may be able to learn the identity of the sender, and the clerks at the last office may be able to learn the identity of the recipient, but nobody can learn the identity of both the sender and the recipient.)

- By mixing the parcels that arrive with those already waiting to be delivered and by filling delivery trucks with empty parcels when there are not sufficient parcels waiting to be delivered, SPS ensures that parcels cannot be traced simply by observing the traffic into and out of their branch offices.

Mix networks are very similar to our hypothetical parcel delivery service. Instead of locking parcels in metal boxes, senders encrypt their communications using the public keys of each mix on the route. The mixes store the messages they receive and at designated intervals randomly forward a message to its destination. If no message is waiting to be sent, the mix randomly generates a message to send. Mixes can also pad messages so that every message in the network is the same length (in our parcel delivery service example people might be able to gain information by observing that lighter parcels are generally closer to their final recipient than heavier parcels; requiring all messages to be the same length in the mix network avoids this problem). Unlike anonymizing proxies, mix networks provide anonymity without requiring a trusted party to forward requests.
Onion Routing

Developed by a group of researchers at the Naval Research Lab, Onion Routing [5] is a system for anonymous and private Internet connections based on mix networks. An Onion Routing user creates a layered data structure called an \textit{onion} that specifies the encryption algorithms and keys to be used as data is transported to the intended recipient. As the data passes through each onion router along the way, one layer of encryption is removed according to the recipe contained in the onion. The request arrives at the recipient in plain text, with only the IP address of the last onion-router on the path. The Naval Research Lab runs a test bed Onion Routing Network that is available for anyone to use. The Onion Routing web site has instructions for configuring a web browser to use this network.

Freedom

A Canadian company called Zero-Knowledge Systems [6] introduced their \textit{Freedom} anonymity system in 1999. Freedom supports anonymous web surfing and email, as well as other anonymous services. Freedom is similar to Onion Routing; however, it is implemented at the IP level rather than the application level. The Freedom client software supports a variety of protocols including HTTP, HTTPS (secure web), SMTP, POP3 (another email protocol), Telnet, IRC (chat), USENET (news), and SSH (a secure tunneling protocol).

Crowds

Developed by researchers at AT&T Labs, Crowds [7] is an anonymity system based on the idea that people can be anonymous when they blend into a crowd. As with mix networks, Crowds users need not trust a single third party in order to maintain their anonymity. Crowds users submit their requests through a \textit{crowd}, a group of web surfers running the Crowds software. Crowds users forward HTTP requests to a randomly-selected member of their Crowd. Neither the end server nor any of the crowd members can determine where the request originated. The main difference between a mix network and the Crowds is in the way paths are determined and packets are encrypted. In mix networks, packets are encrypted according to a pre-determined path before they are submitted to the network; in Crowds, a path is configured as a request traverses the network and each crowd member encrypts the request for the next member on the path. One advantage of the Crowds approach over a mix network is that it is much easier to dynamically reconfigure paths when the network changes. A disadvantage is that Crowds relies on the computer systems used by Crowds members rather than dedicated mix computers, and thus added latencies may be introduced. However, Crowds has been optimized for efficiency. Because it does not use public key cryptography it can be implemented more efficiently than most mix networks. Crowds has been implemented only for the HTTP protocol, but it could be extended easily to support other protocols.

Anonymous Email

A variety of services are available for sending anonymous email. These services offer
varying degrees of anonymity and a number of different types of interfaces [8].

Anonymous remailers are similar to anonymizing proxies: they strip off identifying information from email headers and forward the email message to its destination. Your email address will be unknown to the recipient of your email; however, it may not be unknown to the remailer operator. If you use a remailer for an illegal activity, the operator may be subpoenaed to reveal your email address. Some remailer operators have decided to shut down their services rather than be forced to respond to such subpoenas. To better protect their anonymity, remailer users often encrypt their messages and send them through a chain of remailers. Only the first remailer in the chain knows the userís real email address. And users can take steps to hide their address even from the first remailer.

Most remailers support a variety of features including the ability to accept encrypted messages that contain instructions for processing and the ability to send replies back to an anonymous sender.

**Cypherpunk remailers**

Cypherpunk (also called Type-I) remailers are generally pretty easy to use, even without special software. Web interfaces are the easiest way to use a Cypherpunk remailer. Users can also use standard email clients to send email to a remailer with extra headers containing instructions for the remailer. In addition, utilities are available that simplify using anonymous remailers and encryption programs [9].

Cypherpunk remailers are vulnerable to a variety of attacks, due mostly to the fact that someone watching the traffic into and out of the remailer can often link incoming and outgoing messages based on their size. This is possible even if the remailer holds messages until other messages arrive and releases them in a random order. Occasional remailer users probably do not need to worry too much about this, but these vulnerabilities may be of concern to frequent remailer users.

**Mixmaster remailers**

Mixmaster (also called Type-II) remailers are resistant to most attacks due to the fact that they require all email messages to be broken up into fixed-sized chunks.[10] Thus, all messages going in and out of a Mixmaster remailer are the same size, and therefore linking incoming and outgoing messages is much more difficult. Because messages have to be specially prepared and split into fixed-sized chunks, Mixmaster remailers require the use of special software.

**Encryption Tools**

Whenever sensitive information is stored on a computer or transmitted over the Internet, it should be protected using strong encryption. A detailed discussion of encryption algorithms and products is beyond the scope of this paper. However we will touch briefly on some basic types of encryption products that are useful for helping individuals protect their privacy.
Encryption algorithms are based on mathematical functions that are easy to compute if you know the secret number contained in the encryption key, but very difficult to compute otherwise. If somebody wants to decrypt a file but they don’t have the key, they can try every possible key until they find the right one. This is called a brute force attack. If the key is not very many digits long, a computer can try every possible key pretty quickly. But if you make the key long enough, it will take even the fastest known computers many years to try every possible key. Thus the probability of discovering the key you used within your lifetime would be very small. As computers have gotten faster, longer keys have been necessary to avoid brute force attacks. Until recently, many people used 56-bit encryption keys. However, there are now super computers that can find 56-bit keys in a matter of days [11]. Thus security experts currently recommend the use of 128-bit keys.

Many encryption tools still use keys that are much shorter than 128 bits. In fact, because of legal restrictions on exporting strong encryption from the United States (lifted in 1999), some commercial software developers use 40-bit encryption. A brute force attack can uncover a 40-bit key in less than a minute (or in a few seconds if you use a super computer). These keys may be sufficient to protect your files from casual snoopers, but they are clearly insufficient for any applications that require real security.

File Encryption

Even if you are not planning on sending files over the Internet, if you are storing sensitive information on your computer you may wish to encrypt it to protect it from nosy co-workers or family members as well intruders (including both electronic and physical intruders). If your computer is connected to the Internet, a virus or worm could potentially copy any of your files and transmit them to others. If you store files on a shared computer, other users and system administrators may be able to gain access to your files. By encrypting your files you prevent them from being read by anyone who does not know the encryption key.

A variety of commercial software packages can be used for file encryption. Be wary of those that come with word processor or spreadsheet programs—they often use very weak encryption algorithms. There are also a number of shareware and free encryption programs, some of which offer good strong encryption.

Pretty Good Privacy (PGP) [12] is one of the best-known encryption programs. It was originally developed by Philip Zimmermann and distributed for free.

Email Encryption

As email travels across the Internet it may pass through many points along the way where it may be intercepted and read. Server administrators may read email stored on a server. And some companies routinely scan the email of all their employees (there are even products marketed specifically for this purpose!). You can use encryption to protect the email you send from being read by anyone other than the intended recipient.
Encryption capabilities of popular mail programs

Recent versions of many popular email programs include the ability to send and receive encrypted email and to sign email messages with digital signatures. Microsoft Outlook, Microsoft Outlook Express, and Netscape Messenger can all send signed messages from users who have obtained a digital certificate such as a VeriSign Digital ID [13]. They can also be used to send encrypted messages to other users who have such a certificate. Digital certificates can often be obtained free of charge for a trial period, but after that users are generally required to pay a small annual fee. Other program such as Eudora light and Eudora Pro are available with PGP built in. Because users can generate their own PGP keys, users of these programs need not purchase digital certificates.

Plug-in programs are available that add PGP support to other email programs including Microsoft Outlook, Microsoft Outlook Express, Microsoft Exchange, Lotus Notes, Netscape Messenger, Pegasus Mail, Claris Emailer, Elm, Pine, Zmail, and Emacs.

Web-based encrypted email

Many people do not use email programs on their own computers, but instead rely on free web-based email services that allow them to access their email from any web browser. A number of new services offer web-based encrypted email [14]. These services use a variety of techniques for encrypting email, including providing java applet encryption programs and SSL connections established by a web browser (see the next section for information about SSL). These services are convenient because they do not require users to download special software or purchase digital certificates. However, most are much more vulnerable to attack than other types of encrypted email [15].

Another interesting service, offered by Disappearing, Inc. [16], allows people to send encrypted email that effectively self-destructs at a time indicated by the sender. The system works by requiring recipients to fetch special keys from the Disappearing, Inc. server in order to decrypt their email. The keys are destroyed at the specified time and the email can no longer be decrypted. Of course, if the sender or recipient save decrypted copies of the message it will not disappear. But this system does prevent old email from being retrieved by nosey system administrators or as a result of court order to submit email relevant to a particular case.

Encrypted Network Connections

Email and file encryption are useful for protection your data in storage and when it is being transmitted via email, but additional steps are necessary to protect the communications between your computer and other computers on the Internet. If you fill out forms on web sites or use programs such as telnet to login to remote computers, you may be sending sensitive data across the Internet. In this section we will introduce two protocols for transmitting encrypted data over the Internet. In addition, there is a variety of software on the market that allows users to create secure tunnels between their computer and other computers on the Internet or to establish virtual private networks that allow data exchanged between computers to remain private.
traveling on a corporate intranet—even though it is really traveling on the public Internet.

**Secure Socket Layer**

The Secure Socket Layer (SSL) is a general-purpose protocol for transmitting encrypted data over the Internet. It is used by all of the major web browsers for securely transmitting data to web sites that support SSL. The SSL protocol provides for encrypted data transmission, as well as authentication of clients and servers. Web sites that use SSL transmit digital certificates that browsers can use to authenticate them.

When you use your web browser to fill out online forms, it is a good idea to check to see whether SSL is being used. Without SSL, the data you submit will be sent across the Internet in plain text for all to see. Internet Explorer displays a closed pad lock icon in the bottom right-hand corner of its window when SSL is in use. Netscape Navigator 4 displays a closed pad lock icon in the bottom left-hand corner of its window to indicate that SSL is in use and an open pad lock icon to indicate that SSL is not being used. (Previous versions of Netscape Navigator displayed keys and broken keys.) Clicking on the Netscape lock icon or double-clicking on the Internet Explorer lock icon bring up additional security information.

Even if a web site uses SSL, your data may not be well protected if the site does not use a strong encryption algorithm. Some versions of web browsers use weak 40-bit encryption. However, 128-bit strong encryption versions are also available. And some financial web sites now require visitors to have a browser capable of 128-bit encryption in order to access their services. So to better protect your data and to take advantage of these secure financial services, download a strong encryption version of your favorite web browser.

Once you have a web browser capable of 128-bit encryption, you may want to check to see if the sites you visit actually use strong encryption. After clicking on the Netscape lock icon, you can click on the ‘Open Page Info’ button and find out the strength of the encryption the web site you are visiting is using. You can get this information from Internet Explorer by hovering your mouse over the lock icon.

Both Netscape and Internet Explorer also have the ability to warn users about potentially dangerous situations such as when they enter and leave web sites secured with SSL, when they are about to submit unencrypted forms, and when a web site sends an invalid certificate. Users can control whether or not warning are displayed in each of several situations.

**Secure Shell**

Secure Shell (SSH) [17] is a program that allows users to log into other computers on a network using secure variations on standard Unix commands. It also allows for secure X Window System connections and secure forwarding of TCP connections such as HTTP and POP connections.

If you use a Unix system, use the SSH replacements for rcp to ensure that your passwords and data will be transmitted securely. If you frequently open X connections
over the Internet, consider using SSH to establish secure X connections as well. Once SSH is configured on your system, SSH will be invoked automatically whenever you open an X connection and all data transmitted over X connections will be secured.

If you frequently use a Windows system to connect over the Internet to other systems (for example to a computer at your office), SSH may also be useful for establishing secure telnet sessions, accessing a corporate Intranet via a web browser, or securely accessing POP mail servers. However, you might also want to consider virtual private network systems such as TimeStep (http://www.timestep.com/) that offer more comprehensive solutions.

**Filters**

Filters are tools that selectively block email messages, web pages, news groups, HTML cookies, HTML headers, specific words contained in web pages or email, or other content. Filters may be used by parents to prevent their children from accessing adult content, or by individuals to block unwanted email. They may be used to prevent web sites from sending cookies, or to prevent children from providing personal information to strangers. Our focus here is on using filters primarily to help protect privacy.

**Cookie Cutters**

Cookie cutters are utility programs that prevent your computer from exchanging cookies with web sites. Some cookie cutters block all cookies, others can be configured to selectively block certain cookies, and still others remove the cookies after the fact. Some cookie cutters, also remove web site banner ads and prevent your browser from sending identifying header information to web sites. Junkbusters maintains a list of anti-cookie measures [18].

**Child Protection Software**

Adults should know better than to provide personally identifying information to strangers they meet in chat rooms, but children may not. Some parents use filtering software to prevent their kids from typing in their name and home phone number (or parentis credit card number) and sending it to web sites or chat rooms. They may also subscribe to services that restrict the email addresses with which their children may correspond.

Many child-protection software tools and kidsí Internet services have features that parents can use to help protect a childís privacy. For example, America Online Parental Controls allow parents to configure a childís account to allow email to be sent only to a pre-approved list of recipients, and to restrict a childís access to chat rooms and instant messaging. Many software tools prevent a child from transmitting certain personal information (specified by a parent) over the Internet. The GetNetWise web site [19] has an extensive list of child protection software. You can search this site for tools that have privacy-protection features built in.
Platform for Privacy Preferences

The Platform for Privacy Preferences (P3P) [20], is a framework for informed online interactions. It provides a simple way for web sites to convey their privacy policies in a way that users’ browsers can read. P3P-enabled browsers can then compare a site’s privacy policy with a user’s preferences. P3P was developed by the World Wide Web Consortium.

For P3P to be successful it must be implemented on web sites, as well as in user agent software. Web sites can implement P3P by translating their human-readable privacy policies into machine-readable P3P policies (encoded in an XML format) and configuring their servers to advertise the location of their P3P policies. P3P user agent software can be implemented in a web browser, plug-in, or other application. P3P user agent software should be able to automatically check for P3P policies at web sites and inform the user of a site’s information practices (in both machine- and human-readable formats). Thus users need not read the privacy policies at every site they visit. P3P user agent implementations are currently under development.

P3P is complimentary to other privacy tools, providing the most assistance to users when they visit web sites where they want to provide information necessary to receive a service but they want to know how that information will be used.

Identity Management Tools

Since the beginning of 1999, at least a dozen companies have announced new services and tools that help people manage their online identities and protect their privacy [21]. Most of these tools allow users to store information in secure personal data stores and use it in conjunction with automatic form filling features. Some restrict automatic form filling to sites that have policies that match a user’s privacy preferences. Some have mechanisms that allow users to opt-in to automatically sharing information with marketers or products or services they have expressed interest in; sometimes anonymously; sometimes in exchange for discounts, coupons, or monetary compensation.

Other Tools

Every week, new online privacy-related tools seem to emerge on the market. Here are a few that do not fall readily into the categories of tools discussed here.

Window Washer [22] is a utility that automatically cleans up the files left behind on your computer by web browsers and email programs. It removes all traces of what web sites you visited, what files you viewed, and what files you deleted. Besides helping you protect your privacy from others with whom you may share a computer, removing these files frees up disk space and often improves system performance.

Topclick [23] is a search engine that is free of cookies or banner ads. The web site also includes a variety of privacy-related information.
Conclusions

This paper has provided an overview of currently available privacy tools. New privacy software and services are being developed every day.

Finally, a note of caution. When selecting security and privacy software tools and services, be careful. Many companies offer security tools that do not live up to their marketing claims. If it sounds too good to be true, it just might be. Be especially suspicious of tools that claim to use proprietary encryption algorithms or that guarantee absolute unbreakable security. Generally the most reliable encryption algorithms are those that have been published and have undergone years of public scrutiny. Likewise open source tools can be more easily analyzed by security experts for possible problems. Look for reviews of privacy and security products by reputable security experts.

End Notes

[1] The Anonymizer (http://www.anonymizer.com/) is a commercial service that offers both fee-based services and a free service supported by advertising.

[2] Privada (http://www.privada.net/)


[8] Electronic Frontiers Georgia maintains a list of anonymous remailer services and related information and tools at http://anon.efga.org/~rlist/


[10] Lance Cottrell describes the vulnerabilities of cypherpunk remailers and the design of Mixmaster remailers in an essay at http://www.obscura.com/~loki/remailer-essay.html. See the Mixmaster FAQ (http://www.obscura.com/~loki/remailer/mixmaster-faq.html) for information about obtaining Mixmaster software. In addition Private Idaho and other utilities can be used to prepare messages for a Mixmaster remailer, and web interfaces are now available for Mixmaster.

computer built by the Electronic Frontier Foundation searched 88 billion keys every second and found a 56-bit key in 56 hours. In 1999 the DES Cracker was used in conjunction with a worldwide network PCs to find a 56-bit key in just 22 hours. For more information about the DES Cracker see http://www.eff.org/DEScracker/.

[12] The PGPi project home page has pointers to the latest free version (http://www.pgpi.org/) as well as to a variety of PGP-related products. A commercial version is now available from Network Associates (http://www.nai.com/). A variety of utilities are available to make PGP easier to use for common tasks. For example PGPdisk allows you to create encrypted disk partitions.


[14] Web based encrypted email services include Hushmail, YNNmail, Ziplip, and ZixMail.


[16] Disappearing, Inc. (http://www.disappearing.com/)

[17] SSH can be used freely for non-commercial purposes and by companies for certain internal purposes. The SSH FAQ (http://www.ssh.org/faq.html) contains additional information about obtaining and using SSH. Commercial versions of SSH can be obtained from Data Fellows (http://www.datafellows.com/) and SSH Communications Security (http://www.ssh.org/).


[20] P3P (http://www.w3.org/P3P/)


[23] Topclick (http://www.topclick.com/)

[24] Most software is distributed to end users in a binary format that is ready for a computer to run. However, it is difficult or impossible for a programmer to look at the binary code and gain much of an understanding of how the software works. To gain such an understanding, programmers use the original source code for the software, usually written in a high-level programming language such as C, C++, or Java. The term
open source refers to software for which source code is distributed with a license that allows programmers to not only read the source code, but also modify it and incorporate it into other software. For more information about open source see http://www.opensource.org/.