The Privacy Practices of Web Browser Extensions

PRIVACY FOUNDATION

DAVID M. MARTIN JR., RICHARD M. SMITH,
MICHAEL BRITTAIN, IVAN FETCH, HAILIN WU

DECEMBER 6, 2000

ABSTRACT

Computer users are spending more time with their Web browsers, due to both improved browser technology and industry efforts to provide first-class support through Web interfaces. Users are therefore motivated to customize and otherwise improve their Web browser experiences through third-party browser extensions; at the same time, these desired extensions are able to monitor and report on users' behavior within their primary Internet interface. In this article we report on the privacy practices of some common Internet Explorer extensions. We find examples of both very good and very bad practices and recommend strategies for respecting privacy in this sensitive area.

This work is supported by the Privacy Foundation
http://www.privacyfoundation.org
and the University of Denver.

I. INTRODUCTION

Do today's commercial free-of-charge software downloads show appropriate respect for the privacy of the computer user? In this article*, we explain why we think the answer is "no".

Our research group at the University of Denver Privacy Center addressed the question by investigating Internet Explorer browser extensions. These downloadable pieces of software improve Internet Explorer by giving it the ability to automatically fill out Web forms, perform price comparisons while shopping on-line, or maybe just liven up the interface with thematic images and sounds. A recent summary of browser extension products is available in [1].

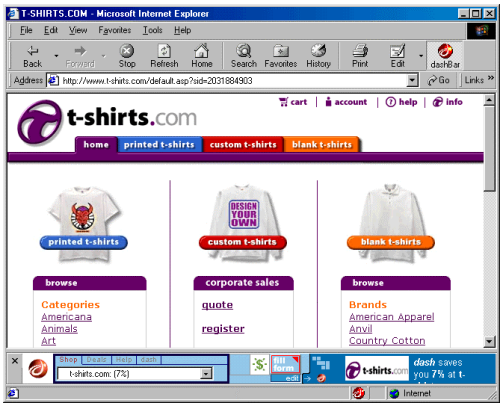


Figure 1: Internet Explorer with a browser extension bar shown at bottom

Browser extensions are usually free of charge — in exchange for clickstream and profile information about the user and access to the user's display for advertising. The spectrum of information practices among products is very broad. While some require full user address information and formal registration, others only request a zip code or age range at download time. Some products create a complete record of every Web site the user visits (e.g., for targeted marketing purposes), while other products avoid performing any actions that could leave an audit trail hinting at the user's Internet activities and personal interests.

Whatever they are, the full terms of exchange are seldom made clear to users. Although advertising components and requests for personal information are apparent, tracking a user's Internet activities is an inherently invisible act. Users enticed to download software described as "totally FREE" have no *a priori* reason to suspect that the software will report on their Internet usage.

The difference between "free gift" and "free of charge" is significant. A gift is given with no expectation of compensation. A haircut at the local beauty school may be free of charge, but it is no gift: the school extracts training value from the exercise, and the customer suffers increased risks. Public domain software is usually a gift. Software that tracks users for business purposes certainly isn't. Vendors and users must understand that tracking records are potential subpoena and mining targets as long as they exist, whether in primary, backup, system log, or debug form.

A. SUMMARY OF THE RESULTS

We downloaded 16 Internet Explorer browser extensions and watched them at work. A number were well behaved. But others seemed to outright exploit our hospitality, watching and reporting

* An abridged version of this article is to appear in the February 2001 issue of *Communications of the ACM*.

our every move in the browser, some intercepting data sent to competitors and others reporting back to headquarters on pages that we "securely" downloaded using Secure Socket Layers. We tried to reconcile the observed behavior with the corresponding software's privacy policy and license agreement and noted any discrepancies. We then compared the products against the well-established standards of Fair Information Practices [2].

We found that 100% of the tested products had significant problems disclosing their behavior to end users; more than 50% of the products employed data flow that appears to weaken end users' privacy, even though the product functionality marketed to users did not require it; and 19% (3 of the 16) products are no longer available or have substantially changed since we first reviewed them.

Although we are disappointed with the state of the art, we recognize that the Internet is a very new deployment arena and missteps are to be expected. By pointing out problem areas and highlighting good practices, we hope to constructively contribute to the debate shaping Web privacy standards.

B. OUTLINE OF THIS REPORT

After the introduction, we set the stage by discussing privacy expectations, fair information practices, and historical cases in part II. We describe our methodology and we present our general findings in part III. The specific product findings appear in part IV, and we conclude in part V. Parts VI and VII contain our lab notes and resources for further study.

TABLE OF CONTENTS

II. BACKGROUND

A. INFORMATION AND TECHNOLOGIES IN WEB PRIVACY

Most computer security threats are best imagined with a time frame in mind:

- "The intruder can log on to the computer until the operating system is reinstalled"
- "The April Denial-of-Service attack lasted for almost 5 hours"
- "This compromised key can be used to sign documents on my behalf until notice of the key revocation reaches the ends of the network or the key expires"

Breaches of privacy, however, are permanent. We know of no practical way to revoke knowledge once gained; we can only hope that knowledge inappropriately gained will not be inappropriately used. In this section we describe some of the information types and technologies that we look at in a privacy investigation of Web-based software.

1. PII: PERSONALLY IDENTIFIABLE INFORMATION

PII includes data such as a person's name, street address, telephone number, email address, social security number, etc. Clearly, this is among the most sensitive type of information to consider.

2. NON-PII

This strange term is used in the Internet marketing community to denote any and all information gathered about a user as long as it is not directly associated with PII. For example, a program that reports only the host name of the web sites a user visits but never identifies the user directly would probably be said to monitor "only non-PII". The intent of the gatherer matters a great deal here: although an expert may be able to identify a user by examining enough non-PII traces, the process would be very difficult to automate.

3. IP ADDRESSES

IP addresses identify computers and not necessarily users, but the correspondence is sometimes straightforward. An IP address usually suggests at least an organizational affiliation. In many circumstances such as home and laptop use, IP addresses are assigned dynamically and therefore somewhat unpredictably. But even then, audit logs of address assignments are routinely archived. In combination with caller-ID logs maintained by Internet service providers, an adversary armed with a search warrant or a bribe could quite reasonably trace a dynamic IP address down to a few square meters of the Earth's surface. To the extent that there is no central registry matching IP addresses to user identities, IP addresses are often regarded as non-PII in official privacy statements, but this position is clearly debatable.

4. COOKIES

Also known as "client-side state", cookies are the mechanism by which a Web site can register and later retrieve or modify a modest amount of information with each Web browser that contacts the site. In the most privacy-threatening application of cookies, a site can invent and assign a unique pseudonym (usually a number) to each Web browser it encounters. The Web site, which has no straightforward way to recognize the Web browser in advance, chooses the identifiers without association to PII. Since there is generally a one-to-one correspondence between Web browsers and their users, this technique enables the site to keep full information about its

interactions with each Web browser user when the cookies flow as intended. Still, the Web site has invented its own pseudonym for the user, and there is no reason to expect that the pseudonym would make it easier for the Web site to obtain the user's PII without the user's consent. Most Web browsers allow the user to disable cookies in order to prevent even this low-grade recognition. However, the default in most browsers is to allow cookies to be sent and received without user knowledge or involvement.

5. THIRD-PARTY COOKIES

Cookie exchanges that occur between a Web browser and a third party — not the second party Web site that the browser initially contacts, but an additional site that the second party introduces into the transaction — are third-party cookie exchanges. Third-party cookies are most commonly associated with graphic images embedded in an HTML document. In a typical scenario, the user's browser first requests a Web page from a second-party server. When delivered, the browser discovers that it contains one or more references to images stored on a third-party computer and begins to automatically fetch them. These latter transactions usually include a "Referrer" line containing the URL that embedded the image. Any cookies sent or received during these transactions are third-party cookie transactions.

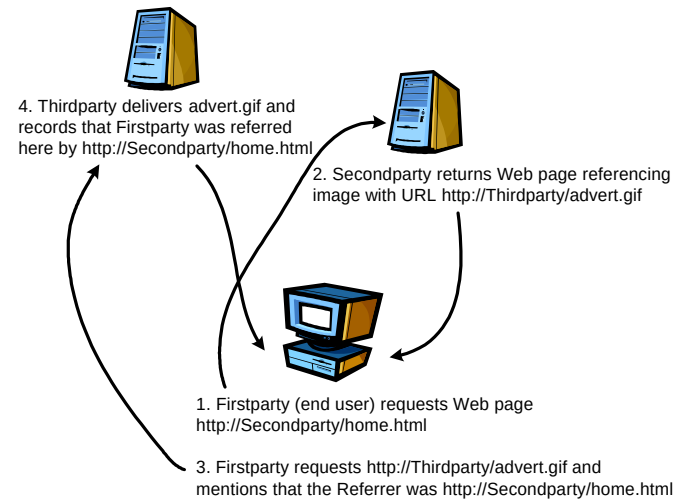


Figure 2: First, second, and third parties

The most operative privacy threat of third-party cookies is that the third party, through systematic logging of the Referrer line, can store records of a user's browsing habits on the pages that specifically call for third-party transactions[*]. Web advertising firms such as DoubleClick, 24/7 Media, and Engage are named as third-parties (for the purpose of fetching advertisements) in much of the Web and are in a position to infer a great deal by examining a user's browsing habits.

Another privacy threat due to third party cookies is at once more subtle and more insidious. If two or more Web sites exchange the pseudonyms that they assigned to a user, then they can also exchange other information about the user. For example, the difference between "Samuel Clemens" and "Mark Twain" is no longer an obstacle in discussion once it is noted that the names are aliases for the same person. Now suppose the second-party Web site knows only the pseudonym "123" for a user, but the third-party Web site knows the person's email address (perhaps because the user once signed up for a mailing list there) in addition to its own pseudonym "456". Once the two sites realize that 123 and 456 denote the same user, the second-party site can request additional information about 123 from the third-party site at its leisure.

[*] Pages viewed that do not explicitly reference the third-party are effectively invisible to the third party.

It is technically straightforward for two such sites to realize they both have data on the same individual. If the URL fetched on the second-party site contains the user's pseudonym[*], then the transaction with the third-party will usually contain a "Referrer" line naming the original URL and therefore the second-party pseudonym. The third-party pseudonym will be carried automatically by the cookie associated with the embedded image transaction. Therefore, this last transaction will contain both pseudonyms simultaneously, enabling the linkage. An interesting property of this data sharing technique is that it need not be envisioned in advance in order to implement it. Since Web sites often automatically log all of the relevant data described above whether they plan to share data or not, a Web site could decide months or years after the actual data collection to begin sharing pseudonyms and only then do the necessary programming.

In general, third-party cookies enable two sites working together to *synchronize cookies*, i.e. to exchange pseudonyms, given sufficient opportunity. As suggested above, PII obtained from one site can then be shared with another as a result. As with ordinary (first-party) cookies, third-party cookie exchanges usually take place without the knowledge or involvement of the user. Note, however, that many sites promise in their privacy policies that they will not engage in this kind of information sharing behavior.

6. UIDs and GUIDs: Unique identifiers

A UID is a pseudonym constructed in order to uniquely identify a user or other entity. The pseudonyms discussed under "cookies" above are types of UIDs. Of particular concern are UIDs that are chosen not by a remote site but rather as a simple function of the user's equipment. For instance, all Ethernet network cards contain a unique Ethernet address (also called a MAC address), and programmers have used these addresses to form UIDs in their applications. One obvious problem with this practice is that logs associating Ethernet addresses with users could be maintained by networking equipment vendors, site administrators, or state authorities, and this information might be sold or subpoenaed. Another problem is that by agreeing in advance to use a user's Ethernet address as part of a UID, no "cookie synchronization"-type step is necessary for cooperating sites to exchange information about the user. In effect, both sites would already be using the same pseudonym for the user — that is, the user's Ethernet address. For these reasons, the industry generally recognizes that UIDs should not contain Ethernet or similar addresses for the best privacy protection.

While the concept of UID is fairly loose, the name "GUID" (globally unique identifier) denotes a specific type of unique identifier that can be automatically constructed at the request of a Win32 program, usually for the purpose of providing unique physical names for COM structures. A program that must uniquely identify a user can just request a new Win32 GUID and use that as the user's pseudonym. Unfortunately, on every Windows platform except Windows Me, the GUIDs generated by the Win32 library contain the computer's Ethernet address if an Ethernet adapter is present. For example, the authors used the program *guidgen* provided with Microsoft Visual Studio to generate the GUID string {F9DFBDA0-C202-11d4-BDCE-00105A9D4FAF}. This string clearly shows the Ethernet address of the machine used to create it: 00105A9D4FAF.

[*] This is often the case; see §III.D. If a pseudonym is not embedded in the URL, the site can still provide the necessary linkage using other techniques.

Instead of relying directly on GUIDs, developers can simply choose a random number from a suitably large space. Standard cryptographic techniques may be applied if more stringent uniqueness guarantees are required.

B. FAIR INFORMATION PRACTICES

Fair information practices (FIP) are a widely recognized set of topics to address when an entity manipulates data about an individual. First described in a 1973 report of the U.S. Department of Health, Education, and Welfare [3], they are defined more carefully for the Internet in a 1998 U.S. Federal Trade Commission report [2].

These fair information practices are not sacrosanct edicts, but rather a starting point for discussion. As evidence we note that there seems to be no good way to build Access mechanisms into many systems, and lively debates continue as to whether uniform guidelines for Choice and Access are even possible to state for Internet software [4].

In our investigation, we compared business' information practices against the FIP standard. We describe the FIPs below, along with our interpretation of how they should apply to downloaded browser extensions. Our product-by-product analysis of success in implementing FIPs is presented in part IV of this report.

NOTICE/AWARENESS

Notice says that individuals should be informed of an entity's information practices such as:

- who collects the information,
- how the information is collected (if it is not obvious),
- how the information will be used,
- who might receive the information in addition to the primary collector,
- whether providing the information is voluntary or required and what happens if an individual refuses to provide it,
- what will happen to the information after the proposed relationship with the entity is terminated, and
- how the collector maintains the security and integrity of the collected information.

Despite profound divergence regarding the practicality of the other FIPs, most practitioners agree that notice is both the most important and the easiest FIP to implement. Without proper notice, users have no opportunity to meaningfully agree to the entity's practices. Fortunately, making notice more accurate does not necessarily require reengineering anything else.

The FTC report [2] states that in the context of a Web page, notice is easily achieved by posting a privacy policy in a conspicuous and unavoidable place. Web sites have taken this advice seriously; today, most sites that contain a privacy policy have it located only one click away from the site's first page. However, an optional Web page disclosure is not ideal for downloaded software (such as a browser extension) that handles information about an individual. In such cases the moment of downloading is a perfect time to explain to the user what the product's information practices are.

Once unavoidably placed, the notice should explain the information practices as clearly and truthfully as possible. Notice should not avoid mention of aspects of the software that may upset some users; those are precisely the issues that users need to know in order to decide whether to proceed. Certainly the last thing that vendors should want is an army of customers who feel misled and abused.

It is important to realize that the consumer and the software vendor may view privacy threats from profoundly different perspectives. Consider a product that sometimes transmits sensitive personal information back to its headquarters due to software architecture or other constraints. Realizing this, the vendor writes "We do not reveal or otherwise act on such information" into their privacy policy. The wary user, however, wonders: what if someone eavesdrops on the network connecting our computers? What if a disgruntled employee abuses the information? What if a stalker or sociopath is hired as the Web administrator? What if a crime is committed involving the vendor's server and it is seized and entered into the public record as evidence? What if a search warrant is issued against the running server? More theatrically, what if such an audit log contradicts a user's sworn testimony? The point is that the vendor's good intent may be moot in the situations most threatening to the user. A vendor's trustworthiness does not imply that the vendor's information architecture is trustworthy[*].

Our guiding principle is that software should only communicate the data necessary to perform the desired function *as described to the user who obtained the software*. This means that software should not monitor and report on a user's computer use unless it is specifically described as monitoring software. Furthermore, explanations of monitoring should never be hidden away in optional side documents (such as most privacy policies). In general, the less the monitoring function itself benefits the consumer, the more prominent the disclosure should be. Proper notice requires effective communication with all users, not just those curious enough to search for it.

In our investigations, we found rampant problems with notice. This is both good news and bad news: good because notice problems can be fixed in many cases by simply updating a privacy policy and making it more prominent, but bad news because poor notice undermines the effectiveness of all of the FIPs. We elaborate on the problems we discovered with notice in the findings section (part IV) of this report.

CHOICE/CONSENT

Once informed, individuals should be given a choice as to whether they agree to the stated practices. This can be accomplished in downloaded software by following the prominent privacy disclosure with a brief dialog asking whether the installation should proceed. All secondary uses of the software — uses beyond those implicit in the request to download — should also be put to user consent. For example, if joining a mailing list is optional, then the user should be given the choice clearly. Preferably, all choices having to do with the service should be accessible on the same screen.

[*] Vendors don't always see it this way. One, for instance, writes "Rather than simply ask you to trust us (as we hope you would have, if we did), we decided to prove our commitment to your privacy," whereupon assurances regarding their treatment of data follow. The implication is that a vendor's word should quell all consumer fears.

The designer of a choice system must decide whether to make each choice "opt-out" or "opt-in". An opt-out arrangement is one in which the user must take explicit action in order to disable the use of information. For instance, a check box reading "Add me to your mailing list" that initially appears checked is an opt-out service, because the user must click on the box in order to not join the mailing list. However, if the check box is unchecked by default, it is an opt-in service — one in which explicit action is required in order to put information to a secondary use. Note that the distinction is predicated on how the information is used and not on how the question is phrased. So, a checkbox reading "Do not add me to your mailing list" that is by default unchecked is still an opt-out list; explicit action is required to avoid the optional mailing list.

It is clear that opt-in designs give users more control than opt-out systems. While we prefer opt-in systems, we find informed opt-out acceptable in some circumstances. After all, if a checkbox is visible, then it is not necessarily an unreasonable burden to ask the user to read the accompanying text and make a choice. However, uninformed opt-out is inexcusable. If users are uninformed either by design or accident, then they will not know there is anything to opt-out of. This simply cannot be characterized as an available choice.

Once downloaded software has been installed, users should be given the choice to stop using the software. Every product we reviewed made this possible, usually through the Windows control panel "Add/Remove Programs" applet. A user should also be able to temporarily disable a monitoring product without fully uninstalling it, but not all of the products we saw made this possible.

The initial decision to download a piece of software is a type of opt-in, and a subsequent decision to remove the software is a type of opt-out. However, this assumes that the user is fully informed. Several of the products we reviewed do not highlight their information practices properly at download time, therefore their users' decisions to download cannot fairly be characterized as consent to the service's information practices.

In particular, if no mechanism exists to permanently erase information gathered about a user should the user decide to stop using the software, then this must be prominently disclosed as part of the initial opt-in installation. Users are accustomed to being able to uninstall software cleanly. A business that creates a permanent relationship with a user's information should not portray their product as an innocent and trivially uninstallable desk accessory.

ACCESS/PARTICIPATION

People move, change their address, and make mistakes. To the extent possible, an entity gathering information about people should provide the opportunity for people to view, contest, correct, or remove information obtained about them that is inaccurate. By way of comparison, U.S. laws guarantee access to individuals' credit records. These records may be examined and contested if they are inaccurate, but users do not have the right to eliminate their credit histories simply because they want to.

Half of the products we reviewed offered little access. With most products, a user's only option was to contact a representative by email and ask for assistance. In some cases where the company was highly motivated to keep a user's PII accurate, an automated interface to update it was provided.

SECURITY/INTEGRITY

A system with poor security can offer at best poor privacy, since security lapses can expose sensitive data. At minimum, sites maintaining personal information should have their practices monitored by a security specialist, and their network transactions should be designed with privacy and data integrity in mind.

Most products we reviewed used the standard username/password approach to ensure that data stored about them is not revealed to the public. In this case it is appropriate to encrypt the username/password network exchange, but not all products do this. Encryption is an important measure because users tend to use the same password in multiple contexts, and one slip can actually expose multiple accounts. This threat is further compounded with products that monitor clickstreams without applying encryption, because an eavesdropper placed anywhere between the user and the clickstream data receiver can learn both the user's password and the sites that the user visits — i.e., the places where the user's password might work. If clickstream data is worth monitoring, it is also worth protecting against eavesdroppers.

Downloaded network-oriented software can also make users vulnerable to new attacks. Many of the products we reviewed include scriptable ActiveX controls, meaning that essentially any HTML source (such as Web pages, email, and Usenet) can invoke the control's methods[*]. Since ActiveX controls have full control over the computer once installed, they must be designed with great care if they are designated scriptable. We did not mount a full-scale investigation of the ActiveX controls included in the products we downloaded, so this remains an area of some concern.

Several products included auto-upgrade mechanisms whereby their software would automatically detect when a new version of software was available. We believe it is important to obtain the user's consent before updating software in this fashion, even if the only choices presented are to upgrade now or discontinue all use. Installing software is an act with security and compatibility implications and should not be undertaken lightly.

ENFORCEMENT/REDRESS

Enforcement is the means to ensure that claimed practices are actual practices. This can be accomplished by enlisting an independent privacy audit, joining an industry association requiring FIP-like standards (such as TRUSTe [18] and BBBOnLine [19]), and/or being subject to civil or criminal penalties for transgressions.

C. LEGAL STANDARDS

In this section we cite some U.S. laws that may apply to browser extensions and other downloaded software. The recurring theme in the laws is that unauthorized uses of a user's computer are unwelcome and legally risky.

[*] In Internet Explorer, scriptable controls can be disabled by changing the computer's security settings for the network "zone" that delivered the document. The default setting for Internet documents is to allow scripting.

ELECTRONIC COMMUNICATIONS PRIVACY ACT

Section 2511 (3) of the ECPA [5] states that electronic communication service providers may not divulge the contents of any communication other than to the intended recipient, unless the originator authorizes the disclosure. If URLs or Web pages can be considered “content” in this context, then it seems particularly important to ensure that a monitoring system gain explicit consent from a user before sharing information. Simply stating this intent in an optional privacy policy may not be enough.

Similarly, 2512 (1) appears to establish penalties for transmitting software “primarily useful for the purpose of the surreptitious interception of wire, oral, or electronic communications.” A browser extension might be considered to surreptitiously monitor Internet use if the software does not obviously require the monitoring for its primary functionality and disclosure of the monitoring is inconspicuous.

Section 2520 allows private citizens to sue for damages when stored electronic information is used without permission. An individual filed such a lawsuit [6] against DoubleClick Inc. in California, alleging that the company was retrieving private information via cookies and selling that information without her consent to companies that were buying advertising space from DoubleClick. The primary complaint was that DoubleClick had an opt-out rather than an opt-in policy for use of private information that it collected from Web consumers.

FEDERAL TRADE COMMISSION ACT

This Act [7] establishes the authority of the Federal Trade Commission to define unfair and deceptive acts and practices in or affecting commerce. As a result, a business practice that contradicts its stated privacy policy may be vulnerable to suit.

The basic consumer protection statute enforced by the FTC is Section 5(a) of the FTC Act, which provides that "unfair or deceptive acts or practices in or affecting commerce are declared unlawful" (15 U.S.C. §45(a)(1)). "Unfair" practices are defined to mean those that "cause[] or [are] likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition" (15 U.S.C. §45(n)). In addition, the FTC enforces [8] a variety of specific consumer protection statutes, including the Children’s Online Privacy Protection Act (COPPA) [9], that prohibit specifically-defined trade practices and specify that violations are to be treated as if they were "unfair or deceptive" acts or practices under Section 5(a).

The FTC prosecuted [10] Geocities for its misleading use of private information collected from people who used its Web site. Some of this data was collected from children. A consent decree was filed in which Geocities agreed to abide by strict privacy rules with regard to user information. Many of the rules Geocities agreed to are now codified in the regulations created under the COPPA.

The European Union has created very strict privacy policies via its Database Privacy Directive that have significant implications for companies in the United States that wish to conduct e-business in Europe, especially given the global nature of the Internet. While the FTC has negotiated a safe-harbor agreement with the EU that will allow individual business to conduct

business with EU consumers if they follow the EU's privacy rules, there are still concerns that the United States does not have standard privacy rules that comply with the EU directive.

CHILDREN'S ONLINE PRIVACY PROTECTION ACT*

This Act [9] protects children's privacy by giving parents the tools to control what information is collected from their children online. Under the Act's implementing Rule (codified at 16 C.F.R. Part 312), operators of commercial Web sites and online services directed to or knowingly collecting personal information from children under 13 must:

(1) notify parents of their information practices;
(2) obtain verifiable parental consent before collecting a child's personal information;
(3) give parents a choice as to whether their child's information will be disclosed to third parties;
(4) provide parents access to their child's information;
(5) let parents prevent further use of collected information;
(6) not require a child to provide more information than is reasonably necessary to participate in an activity; and
(7) maintain the confidentiality, security, and integrity of the information.

In order to encourage active industry self-regulation, the Act also includes a "safe harbor" provision allowing industry groups and others to request FTC approval of self-regulatory guidelines to govern participating Web site compliance with the Rule.

COMPUTER FRAUD AND ABUSE ACT

Although this law [11] most directly addresses the threat of malicious invasion, it also covers uses of a computer that "exceed authorized access." A business whose software accesses a user's computer in an unauthorized manner — such as quietly changing the computer's security settings or transmitting information about the user unexpectedly — may be liable under this act.

SPYWARE CONTROL AND PRIVACY PROTECTION ACT OF 2000

While the laws cited above might present scattershot opportunities to prosecute cases of poor notice and unnecessary data flow in downloaded software, this recently proposed senate bill [12] addresses the issue head-on.

In its October 6 2000 form, the bill characterizes software that collects information about users for the purposes of communicating it as "spyware" and requires that it adhere to clear notice, choice, access, and security practices. In particular, this bill would disallow spyware from disclosing its practices only in an optional privacy policy side document. Violations would be prosecuted as unfair or deceptive acts under the Federal Trade Commission Act.

Many of the browser extensions we reviewed do not appear to conform to the requirements of this bill.

---

* This statement on COPPA is from http://www.ftc.gov/ogc/stat3.htm.

D. HISTORICAL PROBLEMS

Browser extensions are not the only type of software prone to privacy problems. Most software installed on a Windows PC has free reign over that machine — it can observe any user action, alter any setting, and report any finding over the network. Software architects and computer programmers do not always proceed with privacy in mind, as illustrated by the following recent cases:

REALJUKEBOX

This software allows users to play their CDs, transfer CD songs onto their hard drive, provide information about tracks on a CD, etc. In October 1999, Richard M. Smith discovered that RealJukeBox also monitored the music that its users listened to and sent this information back to Real.com's servers along with a GUID identifying the user who registered the software [13]. To make matters worse, it even collected this data when the computer was not connected to the Internet and arranged to transmit it later. This behavior arguably had little to do with the functionality originally offered to consumers. In response to popular outcry, Real.com removed the GUIDs from their software and more carefully disclosed what they were doing.

COMETCURSOR

In order to personalize computers and Web sites, CometCursor lets users choose from a large collection of thematic cursors and temporarily changes the cursor depending on the sites the user visits. For example, a CometCursor user who visits the Garfield comics web site would see the cursor turn into a stylized image of Garfield the cat. In its January 1999 implementation, Richard M. Smith reported that CometCursor also tracked users browsing through member web sites with a GUID [14]. Again, it is unclear why users should expect to leave an audit trail of their Web browsing as a side effect of viewing whimsical cursors. Recent versions of CometCursor have improved privacy practices; their newer UIDs are not based on network addresses, and some versions of CometCursor do not use a UID at all.

RADIATE AND CONDUCENT

Although wholly separate entities, both the Radiate and Conducent companies publish tools and procedures that independent software developers can use to incorporate advertising (and corresponding revenue streams) into their products. The clearest privacy threat in this category is that both modules frequently obtain new advertisements when the computer is connected to the Internet. This means that remote servers could keep track of the times that a computer (and indirectly, its user) is on-line. More dramatic yet wholly unsubstantiated suspicions regarding this software's monitoring of users have been claimed; we mention this only as evidence that some users are extremely uncomfortable with the possibility of monitoring in this category. One reason this topic has received so much attention is that the advertising libraries are permanently installed on computers without user knowledge or consent, and they run independently of the shareware they were bundled with.

ALEXA

A classic clickstream monitor, Alexa essentially *must* report back on a user's URL clickstream in order to provide its advertised service of ratings and other extra information regarding the user's currently viewed web page. However, in January 1999, Richard M. Smith discovered that Alexa

did not remove query strings from URLs before transmitting them back [15]. Query strings can contain sensitive personal information (see §III.D), and since omitting them would not seem to impinge upon Alexa's ability to provide its service, they should have been removed.

III. GENERAL FINDINGS

A. METHODOLOGY: CHOOSING AND OBTAINING THE SOFTWARE

We did not undertake a systematic survey of the market in order to ensure complete coverage. And since browser extensions are designed and constructed for extremely diverse business purposes, we believe that a random sample of products would not meaningfully represent the full spectrum of practices. So, we simply chose to study the products that had been brought to our attention or that we had noticed in our ordinary Web use. The resulting sample corresponds to the marketing effort and success of their creators, at least loosely.

Browser “plug-ins” such as RealPlayer and ShockWave Flash provide additional expressive possibilities such as multimedia streaming and animated presentations, but as with HTML tables or Java applets, their content is typically embedded in a single Web page at a time. Browser extensions, on the other hand, potentially affect every Web transaction. Since plug-ins just add new forms of content and do not alter the Web browser as a whole, we did not include them in our investigation.

The browser extensions we tested were designed for Microsoft Internet Explorer version 5.0. IE is broadly programmable, and many browser extensions are available for it [1]. By trapping certain IE events, it is not hard for a browser extension to observe what the user does within IE and then transmit the audit trail to a remote site. However, we would not characterize the problems we saw as problems with IE; rather, they are misuses of it. Netscape 6 is projected to have extensibility comparable to that of IE.

B. METHODOLOGY: SPYING ON THE SOFTWARE

After evaluating adherence to the FIPs, we turned our attention to data flow, asking the question “Is the product’s data flow consistent with the service offered to the user?”

Uncovering a browser extension's behavior is conceptually a simple task demanding few extraordinary measures. The software’s marketing pitch and privacy policy usually gave us a good idea of what to expect. We then selectively applied tools such as *tcpdump*, *ethereal*, *windump*, a custom logging proxy, a custom TCP stream reassembler, SST Inc.'s *TracePlus32/Web Detective* (which can produce cleartext versions of SSL transactions), *regedit*, the Visual Studio OLE viewer, the disk imaging and recovery tool *Norton Ghost*, and a file viewer for simple cache diving.

Approaching a new product, we used Ghost to save an image of the Windows 98 Second Edition installation before installing the software. Having an OS image allowed us to revert to a “clean” system at will. After enabling an appropriate network monitoring tool and taking an initial snapshot of the registry, we installed the software and ran it while surfing through a locally developed list of test sites and a varying slate of “real-world” sites. Afterwards, we sifted through the piles of accumulated network traces, registry changes, cookies, and other data in order to

understand its data flow architecture. Most products required several iterations of install-surf-interpret before their behavior was clearly exposed. We did not attempt to reverse-engineer compiled code; if we could not answer important questions on our own, we asked the product's manufacturer.

C. THE PROBLEM WITH ENCRYPTION

Some of the browser extensions use encryption and cryptographic authentication in their client/server communications. We approve of this practice, particularly when the communications in question contain sensitive data. Since developers of browser extensions have the ability to control both the client and server ends of the communication, there are few compatibility concerns limiting the use of cryptography.

However, the use of encryption puts privacy researchers and advocates in an awkward position. Having installed network-enabled software onto their own computers, we believe that users or their agents should be able to monitor their computers' communications to ensure that they are behaving as advertised. When we contacted one firm about their encrypted communications, they reported that our inquiry led them to realize that their client had been transmitting the wrong data. Given that ordinary users will not be able to easily decrypt data flowing from their own computers, we wonder how businesses plan to find such errors as encryption becomes more common.

Under the venerable slogan "Trust But Verify", we peeked inside the encrypted form of some exchanges. It is always possible to recover the intended messages by controlling one of the communication endpoints, since each endpoint contains the cleartext at some point in time. In practice, it can involve considerable effort — but nowhere near the effort it would take a third party to mount a brute-force decryption.

It is easy to decrypt SSL (secure socket layer) exchanges. Since SSL is offered as a standard service in Win32, one needs only to tap into the relevant library interface in order to obtain the cleartext. The product TracePlus32/Web Detective from SST Inc. [16] does precisely this. In another case, a product used a custom encryption library, and we used the Microsoft Visual Studio debugger to monitor its transmissions.

We found that all of the communications we decrypted contained more or less what they were supposed to. None of the clients were observed transmitting personal information inappropriately, for instance. Still, we call upon client/server designers who employ encryption to provide a facility for monitoring the information they transmit. We believe that users have the right to see this information. Where possible, we recommend the use of SSL; it is a mature protocol providing secrecy, authentication, integrity checking, and it can be tapped easily at the communication endpoints.

One objection we received to our suggestion is that encryption helps hide the intellectual property embedded in the client/server exchange, and if that were to be revealed, then a firm may lose its competitive advantage. We do not see this as a compelling argument. Since legitimate competitors are precisely those experts sufficiently motivated and sophisticated to unravel the communications (as we did), jealously guarding the cleartext at all times really only protects the communications against unsophisticated end users — who pose no serious competitive threat.

D. INAPPROPRIATE URL MONITORING

The most common problem in the data flow area (in roughly 50% of products reviewed) is the inappropriate monitoring of URLs. URLs can contain a wealth of sensitive parameters. Consider the URL "http://www.google.com/search?q=**AIDS+treatment**&btnG=Google+Search". Clearly, this URL suggests not only that the user does Web searches using Google, but also that the user is interested in AIDS treatment. The part of the URL following the '?' delimiter is called a *query string* and is usually handed over to a program running on the Web server for further processing. At various Web sites we have observed query strings containing usernames, email addresses, physical addresses, telephone numbers, flight numbers, etc. Users have no reason to think that a browser extension that simply tracks "the Web sites you visit" would retransmit this type of information to a remote server, but many extensions do.

The browser extensions themselves are not responsible for personal information appearing in URLs. Nonetheless, their designers should recognize the invasiveness of monitoring this class of URLs and ensure that personal data is stripped away before transmitting the data to remote servers. The easiest approach is simply to strip off all data following any '?' character in URLs before storing or transmitting the URLs to a remote site.

1. http://live.av.com/scripts/search.dll?ep=7&gca=address&orderby=distance&sstreet=**172+mason+terr**&scity=**brookline**&sstate=**MA**&szip=**02446**&scountry=**USA**&query=furniture&qname=&sic=&ck=&ccity=brookline&cstate=MA

   This URL clearly contains an address (we added the bold face). Should this URL be transmitted to a server owned by a company that did not already know this address?

2. http://dps1.travelocity.com/airgdetails.ctl?aln_code=**US**&dep_dt=**19991230**&dep_arp_code=**PHL**&arr_arp_code=**BOS**&flt_num=**2386**&aln_name=US%20Airways&rqs_dow=Thursday&SEQ=946248230535298&last_pgd_page=glblretrieve.pgd

   This URL shows one of the authors confirming his 14-year-old daughter's flight home. Should audit trails like this be stored in perpetuity by dot-coms?

Figure 3: URLs encoding personal information

In addition to monitoring query strings in ordinary HTTP URLs, we have witnessed browser extensions that monitor HTTPS (secure socket layer), FTP, GOPHER, FILE, RES, AND JAVASCRIPT URLs. Monitoring HTTPS URLs is almost an adversarial practice, because vendors design sites (and users visit sites) with HTTPS precisely in order to prevent third parties — such as a browser extension's home company — from viewing or tampering with the data being exchanged. An eavesdropper monitoring the network between a user and an HTTPS site can normally only identify the IP addresses of the computers communicating, not the URL that the user requested. But a browser extension can easily monitor and retransmit the full URL and any sensitive data embedded in it.

Monitoring FTP and the other URL types is inappropriate because these URLs seldom correspond to Web sites in the way that HTTP URLs do. The FILE and RES types, for instance, are normally used to refer to files located on the user's computer. This hardly seems like information that remote sites should be gathering.

Many URL monitoring products do not distinguish between internet and intranet URLs. This means that an intranet URL such as "http://payroll/index.html" will be monitored even though the URL is only meaningful when used within a particular corporate network. Intranet URLs are best ignored by monitoring software.

We believe that many companies actually have no interest in receiving sensitive data and simply did not anticipate this issue when designing their products. For example, the companies whose products we reviewed do not appear to be in the spamming business, so their servers probably just ignore any email addresses embedded in URLs they receive. But by then, the URLs and email addresses have already been inappropriately exposed and possibly logged.

Some browser extensions intentionally examine query strings in the hopes that they will have shopping or profiling value. After noticing a user's Web search for portable MP3 players, for example, a browser extension might present an advertisement or coupon for a specific player. This can be done in a privacy-friendly way by reacting to known keywords on the monitored user's PC — i.e., without transmitting the URLs to a remote site and thereby exposing potentially sensitive data. A little defensive programming would make a big difference here.

E. HARVESTING SEARCH STRINGS

Some of the browser extensions specifically monitor the user's search engine submissions. Technically, this is a special case of query string monitoring. The practice is interesting because search engines such as Yahoo! and AltaVista sell advertising responses to search phrases: in addition to the ordinary search results, a banner advertisement related to the search sometimes appears as well. When a browser extension monitors the search strings too, it essentially rides on the search engine's infrastructure without paying any of the infrastructure maintenance costs, and may dilute the value of advertising purchased directly from the search engine if it displays competing advertisements. When an extension goes further and changes the appearance of the search engine site (for instance, by superimposing its logo on top of the search engine result page), questions of copyright infringement, deceptive advertisement, and fraud come to mind.

F. STAYING IN BUSINESS

At the peak of our study, we had 16 products under investigation. As time elapsed, three of the companies (19%) either discontinued their product or substantially changed their business model, so we removed them from the study. The companies were *Jackpot*, *CrowdBurst*, and *Enonymous.* Each of these products had noticeable user tracking components, and their businesses seem to have relied on monitoring as an essential part of their business plan. This supports the hypothesis that user tracking is a risky business proposition. We wonder what will happen to the user data that these firms have already collected [17].

G. DATA FLOW CASE STUDY: THREE BROWSER EXTENSIONS THAT FILL OUT FORMS

Three of the browser extensions we investigated have the ability to fill out shopping-oriented Web forms. The extensions are *Dash, Gator, and Obongo*. In this section, we investigate the data flow of these competing products in order to illustrate the impact that design decisions can have on privacy. Keep in mind that while we have the luxury of analyzing data flow in isolation, these firms had to balance many business and technical requirements in their designs. In addition, we neglect other facets of privacy (such as the FIPs) in this section. Therefore, we caution against inferring overly broad conclusions about the products or companies discussed below.

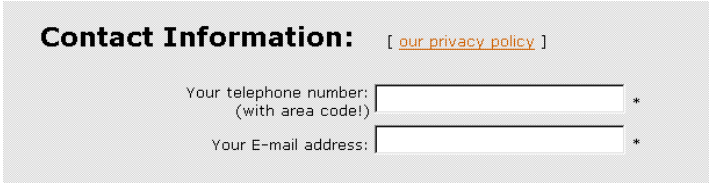


Figure 4: A sample form

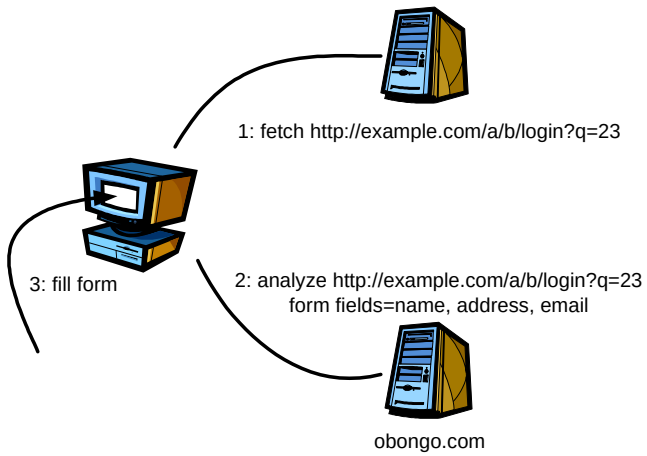
To the user, the products all behave in roughly the same way. After installation, the user configures the software with assorted personally identifiable information (PII) such as name, phone number, credit card number, and username/password pairs for gaining access to restricted Web sites. The software attaches itself to the Web browser and waits until the user reaches a page containing a Web form, at which point the software announces its ability to complete the form for the user. If the user clicks on the software's window appropriately, it will fill out the form fields from the stored PII. For users who do more than a little Web shopping, this is a very welcome piece of automation.

The first design decision is where to store the user's PII. Dash and Gator employ a strong privacy design by storing the PII encrypted on the user's PC only, while Obongo stores it on Obongo's own remote server.

Obongo protects the data from eavesdroppers by encrypting it during network transactions; the main threat is the PII may be mishandled at Obongo. This threat could be soundly mitigated by encrypting the PII on Obongo's server under a key known only to the user, but Obongo does not do this. A representative explained that they did not find it to be a viable option, since users forget their passwords so often. Instead, Obongo relies on business practices and strong internal security procedures to safeguard the data. The advantage of Obongo's remote storage strategy is that a registered user can access the PII from any PC; the other extensions do not support this mobility.

But by maintaining access to the PII, Obongo is in a more precarious position than if they were technically unable to produce the data. For example, a plaintiff pursuing an Obongo user's Hotmail username and password could seek a subpoena asking Obongo to divulge the information. The same action would be absolutely futile against Dash and Gator users, since these companies simply do not have any means to produce the requested information. Note that while ToySmart.com had a privacy policy protecting their customer list, that did not stop them from later trying to sell the list. (Eventually, the U.S. Federal Trade Commission did [17].) In other words, strong technical protections can be far more enduring than guarantees based on legal obligations and business practices.

Another important design decision is where to put the Web form analysis program. In Figure 4, a user observing the positioning of "E-mail address" next to a blank field will conclude that an email address belongs in the field. But form-filling software has to reach the same conclusion by looking only at HTML code. Programming this logic can be challenging because Web designers only ensure that a field's meaning is apparent to humans and not necessarily to programs.



1. User downloads form from example.com and clicks on "fill" button
2. Agent transmits full URL along with page excerpt to Obongo analysis server
3. Agent uses results to fill the form fields on-screen

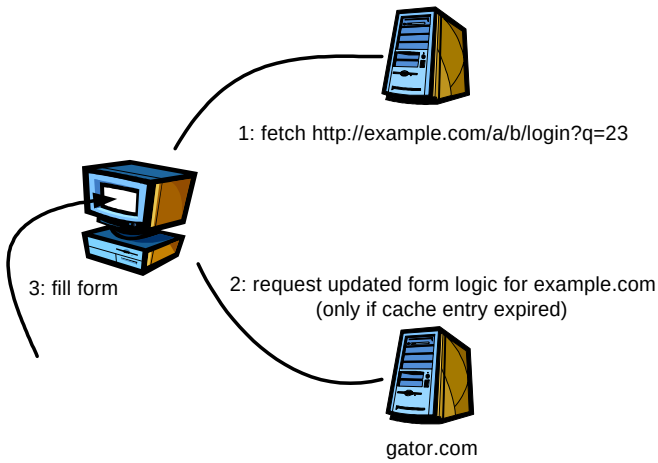
Figure 5: Obongo form analysis

Dash stores the form decoding logic on the local PC, Gator keeps a local cache of remotely supplied logic, and Obongo uses a hybrid local/remote design. By storing the logic on the user's PC, Dash and Gator do not have to rely on a remote server for form field detection, analysis, and completion. (Dash actually does transmit the user's clickstream back to its own server, but not as part of its form filling function.) Keeping the logic local automatically protects those products from the inappropriate URL monitoring problem discussed above. In the case of Obongo, once the user asks it to fill out a form, Obongo sends the URL and small parts of the blank form to its remote server for analysis (see Figure 5). The response to this query will indicate how to complete the form fields from the PII.

Since their server must be consulted whenever an Obongo user requests form completion, the clickstream of completed forms could be stored inappropriately on the Obongo server. But a more dramatic problem is that by sending excerpts from the blank form back to its own server, Obongo rides on the user's credentials to deliver data that the Obongo server may not have been able to obtain on its own. For example, consider a corporate personnel Web site protected by a firewall and authentication dialog. Assuming that these measures adequately protect access, the Web site designers carelessly encode employee names and social security numbers into URLs for their time billing system. The Obongo server would never be able to reach these pages by itself, but a user requesting Obongo's form-filling help would cause the Obongo server to receive some very sensitive information.

A user should not have to fear that software designed to assist the user would transmit excerpts of securely downloaded pages to a remote site, regardless of that site's legal obligations to the user. Although the other form filling extensions also scan and interpret securely downloaded pages,

they do not rely on a remote server for analysis, so they never expose the data. By analogy, few people would want to use a client/server spelling checker with the server based in Redmond, but we all routinely use the spelling checker monolithically contained in Microsoft Word without privacy worries.



1. User downloads form from example.com and clicks on "fill" button
2. Agent requests logic updates for all forms in the domain example.com, if the cached entries have expired
3. Agent applies logic to fill the form fields on-screen

Figure 6: Gator form analysis

Of course, there is a point to Obongo's strategy: recognizing that the form analysis problem is hard, their use of a remote server allows them to adapt to new Web form design styles as they emerge. But the on-line use of a remote server is not the only solution to the problem. For instance, the Gator server sends form-filling logic updates to its extensions when their cached logic falls out of date, and it does so with a mechanism that does not require its extensions to report the user's URL stream (see Figure 6: Gator form analysis). When the Gator extension requests a logic update, it only gives the domain name of the Web site containing the form, which usually contains significantly less information than the full URL. The response contains logic for *all* of the forms on the site.

IV. SPECIFIC FINDINGS

A. PROBLEM AREAS FOR THE REVIEWED PRODUCTS

Table 1 shows the problem areas in the products we reviewed. A product with problems shows either a large or small X, depending on the severity of the problem. A blank indicates that the problem has acceptable (or better) practices in that area. Simply by glancing at the table, it is clear that data flow, notice, and access are widely problematic; there are some problems with security and miscellaneous areas; and choice is implemented pretty well across the board.

Product	Data flow	Notice	Choice	Access	Security	Miscellaneous
@hoc	x	X		x		
AllAdvantage		X		x		
CueCat	X	X			x	
Dash	x	X				
Enfish		X				
Flyswat	X	X		x	X	
Gator	x	X				
iChoose		X	x	x		
NeoPlanet	X	X		x		x
Obongo	X	X	x			X
SurfMonkey	X	X		x	x	
ThirdVoice	X	X		x		
Zack	X	X	X		X	

Table 1: Problems with the reviewed products

7. DATA FLOW

Many products transmit more information about the user than necessary in order to perform the function as marketed to the user. Some also transmit query strings unnecessarily or report on URL or contents of HTTPS or other inappropriate pages.

8. NOTICE, CHOICE, ACCESS, AND SECURITY

Problems with notice are explored in the following section. The other fair information practices are defined in part II of this report. Note that even though poor notice sometimes interfered with the user's ability to exercise choice, we characterized this as a notice problem rather than a choice problem.

9. MISCELLANEOUS

Some products had privacy problems that did not fit neatly into our other categories.

B. DETAILED ACCOUNTING OF PROBLEMS WITH NOTICE

In Table 1, we see that every product had problems with the notice FIP. Table 2 breaks down notice into subcategories. A product with problems shows either a large or small X, depending on the severity of the problem. A blank indicates that the problem has acceptable (or better) practices in that area. At a glance we see that besides a lack of candor — not telling users what it really costs them to use the product — the offenses against notice are widely distributed. The key

to the column headings follows the table, along with the approximate percentage of products with the stated problem.

Product	Lack of candor	Poor placement	Loopholes	Jargon / legalese	Incorrectness	Vagueness	Perspective mismatch	Domain confusion
@hoc	X	X	X	X	x			
AllAdvantage	X	x		x				
CueCat	X		X				X	
Dash	X	X	X	X	X			
Enfish	X	X	X			X		
Flyswat	X	X	x	x	X	X		x
Gator	X			X	x	X		x
iChoose	x	x	x	x	x			
NeoPlanet	X	X	X	X	x	X	X	
Obongo	X	X	X		X	X		X
SurfMonkey	X	X		x				
Thirdvoice	X	x	x		X	X	X	
Zack	X	X		X	X		X	

Table 2: Problems with the Notice FIP

Lack of Candor (100% of products tested). Software that "phones home" more data than is required for the functionality as offered and disclosed to the end user at download time is insufficiently forthcoming. Free-of-charge monitoring products should not masquerade as free gifts.

Poor placement (85%). Products that put their disclosure in a poor location undermine the entire user agreement, since users can't really agree to unknown terms of use. Notice should be unavoidable, particularly in products with unexpected monitoring behavior.

Loopholes (69%). Many sites instruct their users to "return often" to the privacy page to check for policy updates, even when they require an email address as part of the registration procedure, and even when their agents routinely "phone home" for tracking or auto-upgrade purposes. Surely these mechanisms could be used to communicate refinements to privacy statements as well.

Other sites reserve the right to release personal information in order to protect their "property and rights". This loophole could be construed to cover the sale of customer lists due to bankruptcy, appease an aggressive or powerful business partner, or practically any situation in which the company has an economic interest.

During our study we saw approximately 50 privacy policy changes[*]. Most changes were purely cosmetic, but there were significant changes as well. One product (since discontinued) actually changed its policy on sharing personal information: at download time, users had to opt-in to allow their personal information to be shared. However, the privacy policy update retroactively changed this to be an opt-out choice. In other words, users who initially chose not to share their personal information had their preference changed for them; to preserve their original choice, they had to return to the product Web site and indicate their preference a second time! The product manufacturer did send an email to its users containing a link to the new policy, but the email did not summarize the policy change; instead, it included two paragraphs of promotional material.

Technical jargon and legalese (69%). "Clickstream", "URL", "cookie", and "IP address" do not have entries in standard dictionaries, and should be defined if used in a privacy statement. Similarly, putting the privacy statement after 18 paragraphs of legal prose (four of them FULLY CAPITALIZED) makes the statement inscrutable to the average user.

Incorrectness (69%). Sometimes the privacy policy is simply incorrect, and the described behavior is not the behavior we observed. These slips appear to be mostly unintentional, but they do point to a disconnect between policy makers and developers.

Vagueness (46%). A privacy statement that raises more questions than it addresses is probably too vague. For example, one policy states that personal information is held in confidence "except as outlined below," whereupon 17 paragraphs follow, with one of them mentioning the use of the information in order to "make e-commerce faster and easier." Is the personal information held in confidence or not?

Perspective mismatch (31%). A vendor who only addresses data exposure due to a product's ordinary operation within their ordinary business process is only partially addressing a user's concerns. In particular, user data that is transmitted to the vendor must be disclosed even if the vendor does not intentionally store it for long term access.

Domain confusion (23%). Web site privacy seals such as TRUSTe [18] and BBBOnLine [19] cover Web site practices only, but placing these seals near agent descriptions can mislead users into thinking that they apply to the agents downloaded from the site as well. Web sites should not expect users to clearly understand the difference between the browser extension and the Web site, especially when the extension is tightly integrated with the site. In general, we found that sites that do *not* display a privacy seal communicated their policies more effectively, possibly because these sites are not constrained by boilerplate language designed for the Web site context.

Note that while the Platform for Privacy Preferences Project (P3P) [20] specification addresses many of these shortcomings in the context of Web site privacy practices, there is no current specification for communicating the privacy practices of downloaded software to P3P user agents.

[*] Our attempts to automatically detect policy changes using NetMind's Mind-it Web page tracker were complicated by policy pages that include content (such as glowing quotations) that change every time the page is visited.

V. CONCLUSION

We believe that the number of unreported but significant privacy problems in Internet software far exceeds the number of reported privacy-related cases. We also think that most of these unreported problems are best explained by oversight on the part of developers and entrepreneurs unfamiliar with common privacy pitfalls. The Internet is still a relatively new deployment environment, and little guidance is available to those who want to do the right thing. By showing some privacy consequences of early decisions, we hope to help minimize future lapses. Software designers and entrepreneurs may wish to keep our recommendations in mind (see sidebar).

There is a legal sentiment in the U.S.—dating to Justice Harlan's 1967 Supreme Court opinion on wiretaps—stating roughly that privacy protections not codified in law can only exist when society expects them to exist. In other words, by blithely looking away when Web browser extensions surreptitiously send their observations of home life back to headquarters, we run the risk that they will actually accrue that right. It is therefore important to keep society (and professional societies) aware of current privacy practice and developments.

It is time to bring privacy practices in Web software out of the turgid realm of the legal agreement page, elevating them instead to first-class criteria that discerning consumers will count along with speed, memory consumption, and ease of use in the search for the perfect tool for the job.

RECOMMENDATIONS TO SOFTWARE DESIGNERS AND ENTREPRENEURS

1. Tell the truth. If advertisers or others are paying for access to your user's screen or clickstream, don't tell the users that your software is “free”, and don't bury the details of the economic exchange in a legal document. Refer to the ACM Code of Ethics [21].
2. Ask lawyers to review your practices and statements in consultation with engineers familiar with the product’s data flow.
3. Design with privacy in mind; reduce data transmissions to a minimum.
4. Don't force users to opt-out of questionable practices. Instead, invite them to opt-in by explaining how it would help them (or your company). Luring customers into an unwanted situation is not only unethical, but it's also bad for business.
5. If you monitor URLs, choose the URLs that specifically interest you and ignore the rest. In particular, remove text following the query string delimiter ‘?’, and ignore HTTPS, FILE, and MAILTO URLs.
6. If you monitor URLs, design and maintain a mechanism to allow users to access the data collected about them.

VI. LABORATORY NOTES

The following pages comprise our laboratory notes on the products we investigated. We submitted these notes to vendors as an accuracy check, but not all vendors could respond within our publication time frame. In addition, several vendors have expressed their intention to adjust their products or practices in response to our findings.

PRODUCT DESCRIPTION

We installed the version of @hoc that is co-branded with Wired: "The Wired News Toolbar becomes part of your browser. It gives you instant access to the suite of Wired News services from anywhere on the web."

Version: 1.2.16.0, http://www.athoc.com/
Privacy policy: http://www.athoc.com/site/policy.html
End user license agreement: http://www.athoc.com/site/terms.html

DATA FLOW

Over the course of installation, the user submits a name, email address, and a new password to a remote server. Toolbar customizations are stored remotely to allow customizations to appear on multiple computers.

When an option is selected from one of the pull-down menus presented on the @hoc toolbar, a request is sent to the @hoc server containing an encoded representation of the action. The Web browser is then redirected to the appropriate web site corresponding to the chosen toolbar option. Each request to the @hoc server contains a user and a session identifier. An @hoc representative explained that the user identifiers are transmitted in order to enable customization features that were not present at the time of our testing.

A query is periodically sent to www.athoc.com in order to see whether a toolbar update is available. This query contains user and session identifiers and the current toolbar version number.

The system could have instead been designed to periodically download associations between toolbar actions and Web sites to contact, thereby avoiding the tracking behavior we witnessed.

NOTICE

The product is presented as a free gift to the user. The end user license agreement (EULA) represents the product more accurately:

> @hoc will not charge you for this Product. In consideration for the ability to use the Product, you agree to provide certain current, complete, and accurate information about yourself as prompted to do so by the Product registration form and you agree to maintain and update this information as required to keep it current, complete, and accurate...

However, this paragraph is only displayed if the user elects to click on and read the full 16-paragraph EULA.

The privacy policy indicates that the user's name, email address, password, and toolbar customizations will be collected by @hoc. "IP address" is undefined jargon.

The privacy policy states: "Usage Information: @hoc collects usage information such as the web sites and pages you click to, your search terms, and other information related to your web traffic using your toolbar." An @hoc representative clarified that @hoc does not track the user's entire clickstream, but only pages visited as a result of clicking on the toolbar. Search terms which are entered directly into the toolbar's search window are also collected by @hoc. Reading further,

"Technical Information: @hoc collects a variety of technical information such as your IP address, the applications you use, and your operating systems." An @hoc representative reported that the only "application information" collected is the type of web browser (Internet Explorer, Netscape) used and the version number of the browser.

Policy changes
@hoc's privacy policy changed twice during our investigation. (The observations above apply to all versions of the privacy policy we witnessed.) We were not formally notified of the policy changes by @hoc, although one of the changes coincided with questions we had posed regarding their privacy policy, and the new policy was manually sent to us in response to our inquiry.

Policy as of August 8, 2000 ("Beta")
@hoc's privacy policy page begins with a TRUSTe seal and explanation. Although the page does say that TRUSTe only covers information collected through the Web site, most of the statement describes the @hoc toolbar.

It also discloses that "Personal information that you share voluntarily with us may also be combined with data from third parties to improve our services to members." Further down, the privacy policy states that "Only aggregate information is shared with [these] partners and not specific user patterns, unless you instruct or explicitly allow @hoc to share this information." A careful reading suggests that the first statement concerns the Web site only, while the second concerns the toolbar software; however, the privacy policy does not clearly indicate any transition between the two areas of concern.

No indication is given as to how users may be notified if the privacy policy changes.

Policy as of August 15, 2000 ("Beta")
The policy statement now appears to exclusively address the toolbar even though it continues to include the TRUSTe logo and boilerplate text at the beginning. The policy has been reorganized, adopting the nomenclature of fair information practices.

@hoc now states that they "may combine information it collects from you with information from other sources", and acknowledges that they may provide information "to the specific @hoc toolbar Partner who is providing the @hoc toolbar service for your use". Although they typically distribute usage information in aggregated form, they reserve "the right to share additional information with Affiliates."

The policy statement discloses that @hoc may occasionally send its users "vital information" concerning the toolbar software. We never received any such communication during our investigation.

Policy as of August 28, 2000
The policy page no longer bears the word "Beta". All references to TRUSTe have been removed from the privacy page, making it much easier to understand.

Although an email address is still required for registration, @hoc now explicitly reserves the right to change their privacy statement at any time by merely posting the new policy.

CHOICE

The privacy policy indicates: "@hoc, @hoc Partners, and @hoc Affiliates may send you marketing or promotional offers, as well as information concerning usage, tips and tricks, and other editorial content about your toolbar. This can occur via toolbar icons and text, pop-up windows, e-mails and other forms of electronic communications." When clicking on the link indicating how to unsubscribe from receiving such content, the only solution apparent was to uninstall the toolbar.

@hoc states in their privacy policy that uninstalling the toolbar implicitly forbids "future use or disclosure of your information". Apparently, no separate "delete me from the database" request is required to opt-out of the service. However, when we uninstalled the toolbar and reinstalled it later, @hoc recognized us; therefore, uninstallation does not actually delete a user from their database.

ACCESS

@hoc writes that they will consider "reasonable requests" to view a user's usage data, but may charge an administrative fee to do so. We requested to see our own data but received no response from @hoc.

It is possible to alter the registration information you provided during installation of the toolbar, as well as previous customizations that were made to the toolbar.

SECURITY

The @hoc toolbar is installed via an ActiveX control. All transactions to @hoc servers are sent unencrypted.

The privacy policy says that user IDs are generated "randomly", but they appear to be generated sequentially. It appears that session IDs are generated randomly.

Users of the @hoc toolbar submit to having their toolbar automatically upgraded whenever an upgrade is available.

MISCELLANEOUS

The software we tested was labeled "Beta".

RECOMMENDATIONS

1. Make the privacy disclosure more prominent.
2. Disclose why this product is being made available free of charge.
3. Reword the portion of the privacy policy that mentions tracking "the applications you use" in order to prevent undue alarm. @hoc reports that this has been done since this report was prepared.
4. Reconcile language about "random" user IDs with actual practices.
5. Define jargon terms.
6. Provide a mechanism to alert users when the privacy policy changes.
7. Solicit user consent to upgrade when an upgrade becomes available.
8. Use SSL or something comparable to protect communications with the @hoc server.
9. Honor the stated user access policy or make the privacy policy reflect actual practice.

10. Reconcile language forbidding “future use or disclosure of your information” upon toolbar uninstallation with apparent current practice.
11. Provide a mechanism permitting users to terminate their subscription to the service and delete information obtained about them.

PRODUCT DESCRIPTION

"Active AllAdvantage.com members can be paid monthly simply for surfing the Web and receive membership benefits that include sponsor discounts and rebates. In addition, we provide a major incentive for advertisers to build one-on-one relationships with AllAdvantage.com members in an environment of accuracy, privacy and trust."

Version: July 18, 2000, http://www.alladvantage.com
Privacy policy: http://www.alladvantage.com/privacy.asp?refid=

DATA FLOW

Users must enter their name, complete address, and a valid email address in order to register on the AllAdvantage Web site. AllAdvantage uses this information in order to mail payment checks to its users. The Web site also requests the user's phone number, birthday, and general interests.

When visible, the AllAdvantage Viewbar communicates with an advertising server in order to download advertisements, and the AllAdvantage server in order to credit the user's account with the appropriate Web browsing (i.e., advertisement viewing) time.

While the user is surfing, the Viewbar software monitors the URLs visited. Apparently attempting to isolate information that may denote the user's interests rather than which sites were visited, the Viewbar software extracts keywords from the URL and Referrer line. It ignores numbers and the precise formulation of the URL, collapsing a complex URL such as `http://dir.yahoo.com/Health/Diseases_and_Conditions/Prostate_Cancer/` into the keyword string `kw=cancer+prostate+conditions+and+diseases+health;d=yahoo`. It also ignores non-HTTP URLs. After extracting this keyword information, the Viewbar transmits it not to AllAdvantage but, surprisingly, to DoubleClick (aa.doubleclick.net). DoubleClick also provides most of the advertisements to the AllAdvantage Viewbar. Both types of HTTP transactions will usually be accompanied by the user's DoubleClick cookie. Parts of some transmissions to the DoubleClick server appear to be encoded. We did not observe the use of SSL by the AllAdvantage viewbar.

NOTICE

AllAdvantage presents a very long (approx. 55 paragraph) user agreement at registration time. In the second paragraph, they refer readers to the privacy policy, but the nearest link to the privacy policy is at the end of the document, more than 50 paragraphs away. Around paragraph 11, AllAdvantage discloses that they track URLs and keywords. DoubleClick is mentioned neither in the user agreement nor in the privacy policy.

"IP address" and "domain name" are undefined jargon in the privacy policy.

In other respects, AllAdvantage's notice as revealed in their privacy policy is very good. Business practices, data handling practices, and technical limitations are clearly explained. For example, their policy states in layman's terms that while they attempt to filter out potentially personal information in query strings before transmitting, they may not always be fully successful. They draw a distinction between criminal and civil actions seeking subscriber data, guaranteeing that they will notify affected subscribers prior to providing data in civil cases. Instructions to subscribers on exercising choice and accessing data collected about them are easy to understand. Finally, AllAdvantage explicitly addresses the possibility of changes in their privacy practices by

guaranteeing sufficient notice to their subscribers and obtaining their consent before implementing the changes.

CHOICE

As disclosed in the privacy policy, AllAdvantage can be made to stop collecting data by closing (not minimizing) the Viewbar. However, the procedure for closing the Viewbar is nonstandard: the user must right-click on the minimization button and select "close" from a menu. The privacy policy does not give these instructions.

Since users do not know that DoubleClick is serving ads or tracking keywords, the ability to opt-out of DoubleClick's tracking procedure through DoubleClick's Web site opt-out procedure does not amount to a real choice.

The privacy policy gives instructions for canceling a membership and indicates that personal data will be destroyed after their final transactions are complete.

ACCESS

Users may update their personal contact data on-line. However, it is not apparent how to access data collected about a user through the Viewbar.

SECURITY

When a software update becomes available, AllAdvantage requests the user's consent before installing it.

MISCELLANEOUS

In order to encourage enrollment, AllAdvantage rewards subscribers for their referrals. This has led to some spamming problems from overzealous members.

AllAdvantage has recently shifted its emphasis from outright pay-to-surf to sweepstakes opportunities for members, but the Viewbar still appears to be their primary technology component.

RECOMMENDATIONS

1. Disclose the service's monitoring behavior and relationship with DoubleClick.
2. Make the privacy disclosure more prominent.
3. Simplify the user agreement if at all possible.
4. Define jargon terms.
5. Provide a mechanism giving users access to the data gathered about them by the service.
6. Use SSL or something comparable to protect communications between the Viewbar and its servers.

PRODUCT DESCRIPTION

"The :CueCat keystroke automator is a FREE hand-held device that attaches to your computer. About the size of your mouse, the :CueCat reader will change how you use the Internet forever!"

"To use the :CueCat reader, just run its 'nose' across UPC and ISBN codes, proprietary cue codes, and many other standard product codes..."

Version tested: 1.1, http://www.crq.com/crq.html
Privacy policy obtained September 22, 2000: http://www.crq.com/privacy.html

DATA FLOW

Product registration is required, and the user must submit their full name, email address, zip code, gender, and age range to the :CRQ server. A lengthy consumer profile questionnaire is optional. In response to registration, :CRQ transmits a UID "activation code" that uniquely identifies the user. The product manufacturers have stated that the UID is not linkable to personal information in their internal database. (However, see the SECURITY section below.)

Every subsequent use of the :CueCat scanner transmits the information gleaned from the scanned barcode back to the :CRQ server along with the UID and a hardware serial number for the :CueCat device. This gives :CRQ the technical means to compile complete dossiers listing every barcode scanned by each of its users. In addition, by connecting an audio cable to a television receiver as recommended by the manufacturer, the :CRQ software continually samples the sound card in order to interpret "audio cues" — the audio analogs of bar codes. If an audio cue is detected, the :CRQ software transmits its binary representation back to the :CRQ server along with the UID (but no hardware serial number). This gives :CRQ the technical means to track the programming that their subscriber is viewing.

Every use of the :CueCat scanner brings up a navigation bar on the user's screen. Clicking on this bar and manipulating it with the keyboard sends "telemetry" data back to the :CRQ server apparently representing the user's selections.

The :CRQ software does not appear to monitor the user's other Internet activities; only uses of the CueCat bar code reader, interpretation of audio cues, and manipulation of the :CRQ navigation bar result in data flow that can monitor the user's actions.

NOTICE

During installation, :CRQ presents its full license agreement and privacy policy. The software is initially presented as free to the user. The privacy policy hints at the economic exchange by subsequently stating that they may act as a supplier of aggregate information and directed sponsor intermediary. Only on a separate web site (www.digitaldemographics.com) was the intent of the exchange made clear:

> "DigitalDemographics' parallel mission is to gather demographic and psychographic information from our :CRQ users, subscribers, and :CueCat device users...

This separate disclosure indicates that a large amount of data is collected from :CRQ's user base.

After publication of the Privacy Foundation's advisory [22] on the :CueCat and :CRQ software, the above disclosure was removed from the DigitalDemographics Web site. This situation is a clear example of extremely poor placement of notice.

:CRQ clearly states in their own privacy policy that they will not give PII to third parties in order "to solicit you". However, they never disclose that every scan is accompanied by the UID assigned to the user who scanned it.

No indication is given as to how users would be notified if the privacy policy were to change.

CHOICE

:CRQ is inactive until the user starts the application after Windows boots and can be exited in order to disable the software temporarily. No mechanism is apparent to delete information about a user stored in the remote :CRQ database.

ACCESS

An email address is given whereby a user may request copies of the user's "personal profile". No mechanism is apparent for correcting or removing inaccurate profile data.

SECURITY

The registration Web server mistakenly exposed some 140,000 registration records for a short time in early September 2000. These records clearly showed the association between registered users, their UIDs, and their PII. The Web server was quickly fixed.

MISCELLANEOUS

The :CRQ software does not appear to tap into the Web browser like other products in this report, but it does act essentially as a Web accessory.

The :CRQ software consumed 5% of CPU time on our 500 MHz Pentium III when the system appeared to be idle. We believe that it was continuously sampling the audio inputs for audio cues.

RECOMMENDATIONS

1. Disclose why this product is being made available free of charge.
2. Provide a mechanism to alert users when the privacy policy changes.
3. Provide a mechanism giving users access to the data gathered about them by the service.
4. Since the firm promises to never disclose PII, their software has no need to transmit UIDs along with every scan. Send a coarser demographic identifier instead.
5. Provide a mechanism permitting users to terminate their subscription to the service and delete information obtained about them.
6. Use SSL or something comparable to protect communications with the :CRQ server.

Dash

PRODUCT DESCRIPTION

"Links to over 135 leading stores from anywhere on the Web; Hundreds of dollars in instant coupons; Online forms filled out at the touch of a button; The strictest privacy policy on the Web; Up to 20% cash back on every purchase with dash, every day; Special savings alerts directing you to better deals; Credit card fraud protection; Access to weather and search engines without lifting a finger"

Version tested: dashBar 1.4.0.0.26, http://www.dash.com
Privacy policy obtained August 19, 2000: http://www.dash.com/dash/privacy.asp

DATA FLOW

The PII for the form-filling function is stored on the local PC in encrypted form. Startup of the Web browser causes a Dash login transaction. Full HTTP, HTTPS, intranet, and FTP URLs with query strings are all sent to Dash.com. Each transaction with Dash.com carries a cookie identifying the user's name and email address in cleartext.

NOTICE

Dash is presented as free to users, but Dash also extracts value from users by gathering data about their shopping and Web browsing habits and directing users towards affiliated merchants.

Dash's terms of use are presented as optional reading via a hyperlink. Therefore, users may not believe they have agreed to the information practices stated within.

Dash's privacy policy clearly states that they harvest search strings, but they do not mention other types of query strings. URL and IP address are undefined terms. It is unclear whether users will be emailed details if the privacy policy changes.

Dash's claim that cookies are not used for non-members is untrue; we observed the use of session cookies when visiting the Dash site for the first time. The mailing list is opt-out but described as opt-in in the privacy policy.

CHOICE

Users may unenroll from their "My profile" page, in which case all personal information collected about them is removed from the remote database.

Minimizing the Dash bar by clicking on the Dash icon in Internet Explorer makes Dash disappear, but it continues running. Dash can be shut down by right-clicking on the bar or manipulating the Dash icon in the system tray, in which case it stops completely.

ACCESS

User may both view and delete data collected about them (visited sites and search strings) at Dash's Web site. Corrections to data (such as the base value for a cash-back situation) can be requested by email. Dash's implementation of access is very good and demonstrates that providing access to data collected about users can be done within a simple and intuitive interface.

SECURITY

Dash requires the user to enter a password to access their PII every time the Web browser startup, so strangers will find it difficult to access the Dash account or local form-filling data.

MISCELLANEOUS

Dash requires cookies to be enabled for the domain dash.com. For most users this means that cookies have to be enabled at every site in order to use Dash, since current versions of Internet Explorer control cookies in an all-or-nothing fashion.

Dash failed to install with an HTTP/1.0 proxy active.

Our email to privacy@dash.com, listed as the official privacy contact, was returned as an unrecognized address during our investigation. The problem was fixed within a couple of days.

A user has the option of withholding all PII (other than the required email address); in this case, Dash donates earned cash-back to charity.

RECOMMENDATIONS

1. Disclose why this product is being made available free of charge.
2. Make the privacy disclosure more prominent.
3. Define jargon terms.
4. Provide a mechanism to alert users when the privacy policy changes.
5. Reconcile privacy policy with actual cookie and mailing list practices.
6. Remove query strings other than those of specific interest before transmitting URLs to the Dash server.
7. Use SSL or something comparable to protect communications with the Dash server.

PRODUCT DESCRIPTION

"Enfish Onespace, a personal desktop portal, is a unique new blend of online service and downloadable software that enables you to work with both Internet and desktop information simultaneously. This free online service automatically integrates your personalized information such as email, documents and other local files with relevant information from the Internet, enabling you to work with everything you need in one convenient place."

Version tested: August 10, 2000, http://www.Enfish.com/products/download.asp
Privacy Policy obtained August 1, 2000: http://www.Enfish.com/privacy.asp

DATA FLOW

The Enfish client contacts its server when the Web browser is launched and periodically thereafter in order to download news. If the user asks the Enfish client to search for books or CDs, it hands off the request to the Enfish server. If the user creates or views a "personal" or "company" contact page, then the Enfish client asks the Enfish server for information it may know about the person or company (for instance, to focus its headline news more appropriately). The contents of the contact page (such as address, phone number, etc.) do not leave the user's PC.

NOTICE

Product is presented as free to the user. The privacy policy, however, immediately points out that they are paid commissions by partner merchants.

Since Enfish's information practices are described only in the privacy policy accessible through a hyperlink, users may not feel they have agreed to those practices.

The transmission of the names of "personal" and "company" pages is not disclosed. A general discussion of cookies is concluded with the vague disclosure that "Enfish uses cookies to enable us to provide our users with a better experience on our web site." Enfish disclaims responsibility for the privacy policies of third party sites that it frames within its own site; although we have not noticed this in action, this is a sizeable loophole. Enfish clearly expects users to check the privacy policy for updates, even though they have ample opportunity to communicate such changes directly to the affected users.

CHOICE

Enfish provides an email address whereby users may request that their data be removed.

Enfish's entire operation is carried out within a standard application window. Exiting the application causes all Enfish activity to cease until the application is restarted.

ACCESS

No online access of account information is apparent. However, an email address is provided whereby a user can manually request that information be updated or removed from the remote Enfish database.

SECURITY

Communication with the Enfish server is unencrypted.

MISCELLANEOUS

Their claim that “We consider your privacy and security our most valuable asset” leaves us wondering who owns what.

The program crashed frequently on our test platform.

RECOMMENDATIONS

1. Disclose why this product is being made available free of charge.
2. Disclose transmission of names of “personal” and “company” contact pages.
3. Make the privacy disclosure more prominent.
4. Explain the site’s use of cookies more carefully.
5. Provide a mechanism to alert users when the privacy policy changes.
6. Address the issue of third-party content framing more carefully.
7. Use SSL or something comparable to protect communications with the Enfish server.

PRODUCT DESCRIPTION

"Flyswat is a speedy new way to get information, so you save time and avoid Web rage! Flyswat lets you: click on any word on your screen and get a choice of links to related info; go directly to the Web page you need - no more endless sifting; reach the Web's best sites instantly from email, word processing, or any Windows application"

Version tested: v2.1 build 6310, http://www.flyswat.com/
Privacy policy obtained July 28, 2000: http://www.flyswat.com/privacy.html
FAQ: http://www.flyswat.com/faq.html
Service Agreement: http://www.flyswat.com/serviceagreement.html

DATA FLOW

While Flyswat is running in Internet Explorer, keyword requests are made to Flyswat every time the user loads a new page. These requests include the full URL and query string for HTTP requests, as well as a user ID and product ID. A Flyswat representative reported that these IDs are currently used to count the number of users and product installations, but may be used as part of future customization features as well. Query strings and URLs can potentially be mined for personal information and interests, but item 17 of their FAQ states that "Flyswat does not record this information" (referring to query strings).

We observed that data submitted to a remote Web site using the POST method is copied and retransmitted to the Flyswat server as well. Since the great majority of Web sites use the POST method (in favor of the less private GET method) to communicate any sensitive information, Flyswat's policy strikes us as invasive and inappropriate. Data transmitted by the POST method can include names, addresses, passwords, social security numbers, credit card numbers, etc.

By ALT+clicking a word in any Windows application, a request is sent to Flyswat that includes the entire line of text that was clicked, the index of the character that was clicked within the line, the user ID, and the product ID. The other words within the line of text may help to put the keyword in context, but may also leak personal or private information to Flyswat.

Flyswat does not attempt to look up keywords for pages delivered by SSL within a web site. Their FAQ states:

> flyswat will only annotate sites to which it has access while maintaining your privacy. If a page is personalized, requires that you log in to access it, or is on a secure server, flyswat protects your privacy and does not read the page to insert yellow underlined links.

No background monitoring occurs unless the user is browsing the Web and the Flyswat control in Internet Explorer is turned on. No requests or pings are made to Flyswat while the user is inactive.

NOTICE

Flyswat is presented as free to the user. Flyswat's revenue stream is apparently derived from placement of affiliated merchants in the Flyswat search results.

No personal information was required during the installation and registration process.

The only notice given during the installation of the software is the service agreement, which is a legal document and gives limited explanation of the Flyswat privacy policies.

There are no other privacy notices on the download page for the software. Since Flyswat's information practices are so inconspicuous, users may feel that they have not agreed to them.

Flyswat has enhanced privacy features that disable some user tracking features, but no notice is given of this during installation. The enhanced privacy features are only described in the FAQ. Among the possibilities are to "disable GET and POST data submission".

The privacy policy, FAQ, and service agreement contain information about the collection and handling of personal data. "Click-stream" is undefined jargon. The jargon terms GET and POST are also included and described briefly.

The privacy section of the FAQ alludes to the tracking of query strings and their usage for tracking search results. This information is not included in the privacy policy. Neither the privacy policy nor the FAQ clearly states that data communicated by the POST method is captured and retransmitted to the Flyswat server.

The privacy policy does not address the issue of policy changes.

The privacy policy clearly states that TRUSTe does not cover the browser extension software. However, the FAQ only discusses the Flyswat browser extension and also bears the TRUSTe logo at the bottom of the page.

CHOICE

Flyswat automatically starts during Windows startup. An option available from the Flyswat icon in the system tray allows for a user to turn off this behavior.

Flyswat allows a user to opt-out of some tracking mechanisms. These are described in the FAQ.

No mechanism is apparent to delete information gathered about a user from the remote Flyswat database.

ACCESS

There is no apparent mechanism for accessing information stored about a user in the remote Flyswat database.

SECURITY

An auto-update feature is detailed in the FAQ and it is stated that the user must click a "Yes, Upgrade Now" button for the auto-update to install itself.

Flyswat does not appear to use encryption when communicating with its server.

Flyswat's FAQ #2 instructs the user to enable the downloading of unsigned ActiveX controls, effectively weakening the user's security, while asserting that this "will not compromise your privacy or security."

MISCELLANEOUS

We did not witness a program crash or automatic diagnostic report. This product is co-branded with many firms such as MySimon and NBCI.

The practice of changing the appearance of copyrighted Web pages, if only slightly, raises some interesting legal questions.

RECOMMENDATIONS

1. Make the privacy disclosure more prominent.
2. Disclose why this product is being made available free of charge.
3. Incorporate the privacy questions in the FAQ into the main privacy disclosure. The privacy disclosure should be the primary source for privacy questions, not the FAQ.
4. Define jargon terms.
5. Provide a mechanism to alert users when the privacy policy changes.
6. Rework Web site so that the TRUSTe statement does not appear to apply to the browser extension software.
7. Do not retransmit POST data to the Flyswat server.
8. Remove query strings, other than those specifically recognized and selected by the browser extension software, before transmitting URLs to the Flyswat server.
9. Provide a mechanism permitting users to terminate their subscription to the service and delete information obtained about them.
10. Provide a mechanism giving users access to the data gathered about them by the service.
11. Use SSL or something comparable to protect communications with the Flyswat server.
12. Fix the error in the FAQ stating that accepting unsigned ActiveX controls does not weaken a user's security.

PRODUCT DESCRIPTION

Gator assists the user during online transactions by "remembering passwords, filling in forms, and bringing special offers" and can also "target consumers based on site visitation or historical behavior" by matching users with relevant promotions and advertising. This is achieved by displaying advertisements and coupons for online merchants when a user encounters a related or competing merchant's site.

Version tested: 1.5: http://www.gator.com
Privacy policy obtained August 3, 2000: http://www.gator.com/help/privacy.html
End-user license agreement obtained August 3, 2000: http://www.gator.com/help/eula.html

DATA FLOW

Gator's privacy policy states: "Gator.com uses a temporary cookie on its Web site during software installation. Once the installation process is complete, the Gator.com cookie is automatically deleted." Although no cookies from Gator.com were monitored during installation, a DoubleClick cookie is set using a Web bug on the software download page.

Gator builds machine- and user-specific GUIDs during installation and registration, which can contain the Ethernet address of the user's network card.

At startup, the Gator client downloads a hash table listing the Web sites whose forms Gator understands. When a user visits one of these sites, a script file is requested from Gator.com that describes how Gator should auto-fill all of the forms within the site's domain. The Gator client caches these files for 48 hours (value found in registry), after which new script files are downloaded when necessary for a site. This design means that Gator.com receives very little data on the precise Web pages that a user visits. Requests for script files can be triggered by HTTP, HTTPS, GOPHER, and FTP requests.

Users assist Gator.com in learning how to fill out unrecognized forms by dragging their personal data onto form fields. In this process, Gator.com collects user IDs and the full URLs of forms, including query strings. This information is used to rate the popularity of forms that are not in Gator.com's database, and determines which forms should be considered for inclusion. It is unclear why collecting user IDs in this circumstance is necessary.

Advertisements and coupons are downloaded in a similar fashion to script files. Clicking on a coupon/advertisement that is presented to the user causes either a request to the online merchant, or a request to Gator.com, which in turn redirects to the merchant. Neither the user ID nor the machine ID is passed to Gator.com during this transaction.

The program is active while running in the system tray and continues to send and receive information to Gator.com and present coupons and advertisements to the user. The Gator software reports to Gator.com when it is uninstalled.

NOTICE

The software is presented as free to the user. Gator's revenue stream is generated through matching and placement of advertisements and online coupons, sometimes even on the site of a Web site advertiser's competitor.

At installation and registration time, Gator.com discloses that they collect the user's first name, email address, zip code, and country. All other data remains encrypted on the user's PC. This is also disclosed in the end-user license agreement (EULA) as “Basic Contact Information”.

The full EULA contains privacy information about the software and is presented to the user during registration. This is a legal document containing legal terms and technical jargon. Privacy matters are discussed after 18 paragraphs of legal prose. The document states that Gator collects “information on your web usage that remains anonymous to third parties”, but does not clearly define what “web usage” data entails, and it is unclear when and where the information is anonymized. When additional users register on the Gator client, they see only a statement regarding the confidentiality of PII on the user’s PC and not the full EULA.

The privacy statement on the Web site clearly attempts to differentiate between the privacy statement for the Gator client and the Web site itself, but an unsophisticated user may not appreciate the difference. The Web site statement could be mistaken as the condensed version of the “complete Gator software end user license agreement and privacy statement” available through a link.

Gator.com states that it will notify users of changes to the EULA or privacy policy via email, assuming that the user's email address on record is current.

CHOICE

The software can be temporarily disabled using a context menu from the system tray icon.

When the software is uninstalled, the option to remove all user data from the client is presented, and will successfully remove all registry keys that previously stored user, banner ad, and script information.

ACCESS

There is no apparent mechanism for viewing information that is gathered through tracking of Web usage.

SECURITY

User information appears to be encrypted and stored within the Windows registry. Domain names are sent to Gator.com in cleartext, but all other data appears to be encrypted. A user must log in to the Gator client before the encrypted data can be used.

The software appears to support an auto-update, judging from the existence of Gator registry keys named AutoUpdate. There doesn't seem to be any method for turning off the auto-update mechanism. Requests for updates were monitored while using the software, but we did not observe new software releases during our testing, so we don’t know how the auto-update would proceed.

RECOMMENDATIONS

Our recommendations apply to the version of the software that we tested. A Gator representative has informed us that the Gator software and Web site have since been improved in a way that addresses recommendations #1, 3, 7, 8, and 9 below.

1. Make the privacy disclosure more prominent.
2. Disclose why this product is being made available free of charge.
3. Correct the privacy disclosure to describe the use of DoubleClick cookies (not Gator cookies) during the installation procedure.
4. Define legal and jargon terms.
5. Explain the meaning of "web usage" more carefully.
6. Rework Web site slightly so that the TRUSTe statement does not appear to apply to the browser extension.
7. Replace Ethernet portion of GUID with a non-identifying number. Gator has indicated their intent to do this in a future release.
8. Remove query strings and user IDs before transmitting URLs to the Gator server during the form learning process. Gator has indicated their intent to do this in a future release.
9. Provide a mechanism permitting users to terminate their subscription to the service and delete information obtained about them.
10. Provide a mechanism giving users access to the data gathered about them by the service.
11. Solicit user consent to upgrade when an upgrade becomes available.

PRODUCT DESCRIPTION

"The iChoose Savings Alert is FREE software that gives you access to better deals on the things you want to buy -- books, music, movies, toys, hardware & software, consumer electronics, pet supplies, and more (coming soon)."

Version tested: 1.118, http://www.ichoose.com
Privacy policy obtained September 11, 2000: http://www.ichoose.com/support/privacy.jsp

DATA FLOW

The user's name, email address, street address, iChoose username and iChoose password are solicited during installation, encrypted, and transmitted to the iChoose server. Optional information, such as telephone and credit card numbers for online purchases, are encrypted and stored on the local computer only. When the user encounters a Web page recognized by iChoose as representing an impending online purchase, the iChoose client transmits information about the purchase to the iChoose server so that it may hunt for other offers on the same product. A sound and blinking icon are used to alert the user that a better deal has been found.

iChoose does not appear to report on visits to Web sites that it does not recognize as e-commerce sites, nor does it appear to analyze product offers presented to the user with SSL.

NOTICE

The product is presented as free to the user. In fact, the software steers users towards merchants affiliated with iChoose. iChoose's home page contains "selling points" to both end users and merchants; end users may or may not understand the tradeoff.

The privacy policy is presented as optional reading during installation. As a result, users may not believe they have agreed to the information practices stated therein.

The privacy policy is comprehensive and easy to read, although "URL" and "IP address" are undefined. One paragraph overreaches by characterizing IP addresses as "completely anonymous". The possibility of privacy policy changes is not mentioned.

CHOICE

iChoose can be disabled by clicking on its system tray icon. Users may opt-out of the mailing list.

No mechanism is apparent for purging records from iChoose's remote database.

iChoose appears to be affiliated with 24/7 Media. Opting-out of tracking by 24/7 Media requires an extra transaction.

ACCESS

No mechanism is apparent for viewing or correcting user records stored in iChoose's remote database.

SECURITY

Almost all transactions with the iChoose server are encrypted.

MISCELLANEOUS

We asked iChoose personnel to decrypt some of the communications that we saw. In response, they described the type of data that is carried in their encrypted communications in general terms. After decrypting some communications on our own, we saw nothing inconsistent with their explanation or privacy policy.

RECOMMENDATIONS

1. Make the privacy disclosure more prominent.
2. Disclose why this product is being made available free of charge.
3. Define jargon terms.
4. Correct statement characterizing IP addresses as “completely anonymous”.
5. Provide a mechanism to alert users when the privacy policy changes.
6. Disclose relationship with 24/7 Media and explain how to opt-out of tracking by 24/7 Media.
7. Provide a mechanism permitting users to terminate their subscription to the service and delete information obtained about them.
8. Provide a mechanism giving users access to the data gathered about them by the service.

PRODUCT DESCRIPTION

NeoPlanet is an "Internet Desktop" that integrates "a web browser, e-mail client, instant messaging, chat, web directory, search engine, and user-created communities all into a single application." It is essentially a modified form of Internet Explorer.

Version tested: 5.1, build 1517: http://www.neoplanet.com
Privacy policy obtained September 5, 2000:
http://www.neoplanet.com/user_central/privacy/index.html

DATA FLOW

Registration of the software at the NeoPlanet site requires an email address, age, gender, and zip code for the user.

The software sends a login GET request to NeoPlanet servers, opens an SMTP connection to a NeoPlanet server (identified by NeoPlanet support as a diagnostic feature), and attempts to GET an updated version of the software. An option for turning off the auto-update feature in the browser appears in the NeoPlanet control panel, but the software checks for updates even if the setting is turned off.

While running in the system tray, NeoPlanet continues to send GET requests to a NeoPlanet server every 5 minutes. These pings stop when the browser is not in use. The login and ping requests include encrypted strings which contain "usage statistics, system configuration, NeoPlanet ID, and demographic information," according to NeoPlanet's Privacy Policy and support personnel.

NeoPlanet sets a user ID containing the Ethernet address for the PCs network card. This user ID is seen in various requests to NeoPlanet. It is not clearly seen in the encrypted strings that accompany the login and ping requests to NeoPlanet.com, but since the Privacy Policy states that they capture usage statistics along with NeoPlanet IDs, we assume that the same ID is used.

The NeoPlanet ID has been seen in association with some requests to the NeoPlanet Perks features. Another feature in the software is a customization/registration form, used to help the user personalize the NeoPlanet software. When we completed the registration, we saw the NeoPlanet ID (including the Ethernet address) sent to NeoPlanet along with a string of personal information including a full name, email address, marital status, country, zip code, education, birth date, as well as some information polled about usage of the software.

NOTICE

No privacy disclosure was displayed as part of the download process. A link to a privacy policy is present in a sidebar on the NeoPlanet Web site, but it is not clearly related to the download link. The installation displays only the Conditions of Use from NeoPlanet, which is a legal document. The Conditions of Use only provide the address of the Privacy Policy on the NeoPlanet Web site; no link was available. Registration of the product takes place after installation, and presents only a link to the Privacy Policy. The Privacy Policy clearly defines NeoPlanet as an e-commerce and advertising driven company:

> "Our priority to keep NeoPlanet free for our users requires us to provide these advertisers, e-commerce and content partners with aggregate demographic profiles of our user base."

The Privacy Policy details the information that is collected by NeoPlanet, and states that it will only be used/distributed on an aggregate basis. Information collected includes "product usage, system configuration, demographic information and NeoPlanet ID." NeoPlanet collects channel bar usage (the channel bar is a set of links built into the user-interface) and a "predetermined set of URLs." The claim that "Individual surfing information is never collected" is incorrect; in a straightforward example of perspective mismatch, they may have meant to say that their policy is to disassociate surfing information from the individuals as soon as it is collected.

The Policy contains some technical and jargon terms, such as "automated background thread", "aggregate demographic profiles", and "product usage". No information is given regarding notice of updates to the Privacy Policy, even though NeoPlanet does collect email addresses during registration.

CHOICE

A user can temporarily disable NeoPlanet's tracking by exiting the application and using Internet Explorer instead. However, this may not be obvious to all users.

Registration of the software has an opt-out mechanism for product news mailing and promotional materials from NeoPlanet and partners.

No mechanism is apparent to remove information gathered about a user from NeoPlanet's remote database.

ACCESS

Profile information can be changed/removed at any time, although email address, zip code, age and gender must be present.

There is no apparent mechanism for viewing or correcting any user-tracking data related to the use of the NeoPlanet software.

SECURITY

The product supports automatic upgrade, but we did not witness a software upgrade during our testing.

MISCELLANEOUS

NeoPlanet seems to have numerous partnerships with software companies (Flyswat is bundled into the product), content providers (Lycos, and Lycos Communities), and distributors (co-branding of NeoPlanet skins).

Our multiple, specific inquiries regarding "phone-home" behaviors of the NeoPlanet software were met with repeated copies of the privacy policy and no further comment until we contacted the company's CTO directly.

RECOMMENDATIONS

1. Make the privacy disclosure more prominent.
2. Disclose why this product is being made available free of charge.
3. Define jargon and legal terms.
4. Provide a mechanism to alert users when the privacy policy changes.
5. Fix perspective mismatch error stating that "individual surfing information is never collected."
6. Explain the meaning of "web usage" more carefully.
7. Disclose that the software is in communication with the NeoPlanet server as long as the NeoPlanet browser is running.
8. Provide a mechanism permitting users to terminate their subscription to the service and delete information obtained about them.
9. Provide a mechanism giving users access to the data gathered about them by the service.
10. Ensure that the auto-update feature is working as intended, and that user consent is solicited before applying an automatic update.

PRODUCT DESCRIPTION

"Obongo is a free personal Web toolbar that makes your Web experience easy, fast, and secure." Features one-click login and form filling services, Web search, and an email forwarding service.

Version tested August 2, 2000: http://www.Obongo.com/chabi/webSite/index.htm
Privacy policy obtained August 2, 2000:
http://www.Obongo.com/chabi/webSite/privacyIndex.htm

DATA FLOW

Most traffic between the Obongo client and Obongo.com is encrypted with SSL. The user's PII resides on the Obongo server and is transmitted to the PC when the Obongo client starts. When Obongo detects a form or login page, the user has the opportunity to request auto-completion. If selected, the client transmits salient information from the Web page to Obongo.com for analysis, including the full URL (with query strings), Obongo user ID, and excerpts from the page. This information is transmitted even if the page was delivered to the user with SSL. Obongo.com responds by indicating how the client should complete the required information.

If the user directs the Obongo client to perform a price comparison on an item displayed in the Web browser, the client first sends the full URL (including query strings) to Obongo.com. The response redirects the browser to a third party, Clickthebutton.com, which actually performs the comparison and delivers the results.

NOTICE

The software is presented as free to the user. It appears to steer customers towards merchants who are affiliated with Obongo.

Obongo's information practices are presented as optional reading during installation. As a result, users may believe they have not agreed to these practices.

Clickthebutton.com is not mentioned even though it appears to be an essential part of the product architecture. It also appears to fall under the clause stating "no responsibility or liability" for third parties in the Obongo privacy disclosure.

The privacy policy shows TRUSTe and BBBOnLine seals in confusing proximity to discussion of the Obongo software.

Obongo's privacy policy states that personally identifiable information is held in confidence "except as outlined below," whereupon 17 paragraphs follow, with one of them mentioning the use of personal information in order to "make e-commerce faster and easier." Users are warned that additional services, sweepstakes, and promotions may supersede Obongo's privacy policy, but it is unclear whether or how the resultant policies will be accessible to the user after enrollment.

The user's personal address book is said to be stored on the server in a format that "would [not] be retrievable by Obongo", yet we saw it flowing to the server in a form that is cleartext to the server. This is an example of perspective mismatch. Mailing list is opt-out (i.e., the default is to join the list) but mistakenly characterized as opt-in in the privacy policy. Obongo may modify their policy "at any time"; it is unclear whether they will notify users if their policy changes.

CHOICE

Clicking the Obongo icon in Internet Explorer in order to disable it does not actually disable the software, it just hides it.

No mechanism is apparent to remove information stored about a user in the remote Obongo database.

ACCESS

It is possible to change user information online, such as address or credit card information, with the Obongo toolbar.

The privacy policy states that "Users may request to review the information that they have submitted via Obongo's web forms by [clicking here] to contact Customer Service." When we requested this data, we were told that no more information was accessible through this method than data that we can already access through the Obongo toolbar. Specifically, information about a user that contributes to "aggregate or summary information" is not accessible for examination or removal.

SECURITY

The use of SSL effectively removes the threat of eavesdroppers.

MISCELLANEOUS

Obongo creates an email forwarding address for users without warning and allows the user to configure filters that block email from sites at which the user registered using Obongo's assistance. The form-filling feature always supplies this email address, even though Obongo requires a "native" email address for registration. Technically, Obongo may have legal access to the contents or headers of email flowing through their forwarder, as their privacy policy allows them to disclose PII "to the extent necessary to provide [the user] with a requested service", and distributing the email forwarding address is implicitly requesting the forwarding service. We seriously doubt that Obongo or any other company would attempt to rely on such a weak legal technicality to gain access to email contents, but the exposure associated with an extra email hop may not be worth the benefit thrust upon the user.

RECOMMENDATIONS

1. Make the privacy disclosure more prominent.
2. Disclose why this product is being made available free of charge.
3. Disclose the relationship with Clickthebutton.com.
4. Disclose that an @obongo.com email address is automatically created for the user.
5. Explain the product's handling of personal information more precisely.
6. Correct misstatements implying that Obongo cannot access users' personal information and that the mailing list is opt-in.
7. Provide a mechanism to alert users when the privacy policy changes or is overridden by a policy associated with a special promotion.
8. Rework Web site so that the TRUSTe and BBBOnLine seals do not appear to apply to the browser extension.
9. Tighten privacy policy language regarding the confidence of personally identifiable information.

10. Provide a mechanism to temporarily disable the monitoring behavior without completely uninstalling the product. Turning off the Obongo browser bar should probably stop the monitoring.
11. Provide a mechanism permitting users to terminate their subscription to the service and delete information obtained about them.
12. Provide a mechanism giving users access to the clickstream data collected about them by the service.
13. Remove query strings before transmitting URLs to the Obongo server.
14. Do not transmit information about HTTPS URLs to the Obongo server.
15. Provide a mechanism to bypass the use of the @obongo.com email address when filling out Web forms.

PRODUCT DESCRIPTION

"The SurfMonkey Bar is a free navigational tool bar that guides and protects kids during their Internet journeys. Residing at the bottom of your browser window, this protective cyber shield offers comprehensive safety features that enable your children to freely and safely explore the vast resources offered by the Internet. For optimum safety, the Bar uses SurfMonkey.com's proprietary in-page filtering for on-the-fly blocking of inappropriate language, both words and phrases, on Web sites. It also blocks communication with strangers."

Version tested: 1.2, http://www.surfmonkey.com/free_trial/DownloadBar.asp
Privacy policy obtained August 1,2000: http://www.surfmonkey.com/privacy_statement.html

DATA FLOW

To use the SurfMonkey Bar requires setting up an account at the SurfMonkey Web site. Information supplied at sign up time includes an account name, password, real name, email address, and Zip code. Also at least one child's name, birth date, and gender must be supplied. All registration information is associated with the SurfMonkey cookie.

Account information must be verified via a telephone call, US mail, or FAX. If someone registers via telephone, it is technically possible via caller ID to associate a person's phone number with a SurfMonkey account.

When the SurfMonkey Bar is turned on, each time the Web browser goes to a Web page, the first 100 characters of the URL of the page are sent to a SurfMonkey server to check to see if it is a blocked page or not. A GUID is transmitted to the SurfMonkey server with each URL. A new GUID is generated each browser session, but since a GUID usually contains an Ethernet addresses, the GUID still probably identifies a SurfMonkey subscriber.

As an option, the account name can be included with URLs that are sent back to the SurfMonkey servers. This option allows SurfMonkey to track the clickstream of a family member whenever the SurfMonkey Bar is turned on. Why this option is present in the product it is not clear, but it can be turned on remotely without any notice to a user.

The SurfMonkey Bar sends full query strings and HTTPS URLs to the SurfMonkey servers.

NOTICE

The product is presented as "absolutely free" to its users; no notice is given to parents that this product is supported by advertising. "IP address" is undefined.

The SurfMonkey privacy policy does describe accurately what data is being sent to Surf Monkey servers and when. However, the operation of the Surf Monkey Bar is not disclosed to customers on the download page or in the install program. In particular, users are not told that their complete clickstream is sent to SurfMonkey servers. Since this type of exposure is not required to provide the content filtering service, it should be disclosed to users. The license agreement that accompanies the software does not include any references to privacy issues.

SurfMonkey states that if there are material changes in their privacy practices, they will obtain new consent from parents in order to continue use.

CHOICE

No mechanism is apparent for temporarily disabling SurfMonkey.

No mechanism is apparent for deleting information gathered about a user from the remote SurfMonkey database.

ACCESS

It is possible to look at account information and change it at the Surf Monkey Web site. However, no mechanism is apparent for viewing clickstream data collected about a SurfMonkey user.

SECURITY

Account data is stored at the SurfMonkey Web site and is protected by a user-selected password.

A lost password can be retrieved by having it sent to the registered email address associated with an account.

The SurfMonkey Bar is implemented as an ActiveX control that is marked safe for scripting. It is possible to use this control for other Web pages and HTML email messages. For example, we found that an HTML-based email message with the appropriate script code can turn on the tracking feature of the product. It might be possible to maliciously use this control to crash a computer or delete data.

MISCELLANEOUS

If the SurfMonkey servers are not available, it is not possible to surf the Web except by turning off the SurfMonkey bar.

The early release of this product (version 1.0) always included the account name with URLs. Users of version 1.0 of this product should upgrade to version 1.2 to prevent any chance of clickstream data from being associated with their name or email address.

RECOMMENDATIONS

1. Make the privacy disclosure more prominent.
2. Disclose why this product is being made available free of charge.
3. Define jargon terms.
4. Remove query strings before transmitting URLs to the SurfMonkey server.
5. Do not transmit information about HTTPS URLs to the SurfMonkey server.
6. Remove the option allowing an account name to be sent in with all transmissions to the SurfMonkey server.
7. Replace Ethernet portion of GUID with a non-identifying number.
8. Use SSL or something comparable to protect communications with the SurfMonkey server.
9. Ensure that the scriptable ActiveX controls cannot be misused by outsiders.
10. Provide a mechanism giving users access to the data gathered about them by the service.
11. Provide a mechanism in the uninstall program to also terminate their subscription to the service and delete information obtained about them.

PRODUCT DESCRIPTION

"ThirdVoice 2000 is the free personal Web assistant that turns any word on any Web page into links of limitless information." Users can view related topics, discussions, and shopping resources that are linked to any word or Web site.

Version tested: Version 1, build 1.73.2: http://www.thirdvoice.com/
Privacy policy obtained July 26, 2000: http://www.thirdvoice.com/about/privacy.htm

DATA FLOW

The ThirdVoice software sends a login request to a ThirdVoice server when it is started, a logout request when it is shutdown, and pings at about 6 minute intervals while the user is active on the Web.

ThirdVoice associates discussions, research, and shopping links with URLs of Web pages. When a request is sent from Internet Explorer with ThirdVoice running, a request for keywords is also sent out to ThirdVoice. This request includes the user's ID and the entire URL of the page being viewed, including URL query strings. URLs reported to ThirdVoice include HTTP, HTTPS, and even FILE and RES URLs.

Session cookies are used along with these requests in order track a member's activity while using the ThirdVoice software. No persistent cookies were seen while using ThirdVoice.

ThirdVoice automatically starts during Windows startup. Data collection occurs whenever the program is running in the system tray, even if the ThirdVoice user-interface is closed.

Registration only requires a username and email address from the user. Name, gender, year of birth, country, and zip code can voluntarily be entered on a profile page at ThirdVoice.com.

NOTICE

The service is presented as free to users. ThirdVoice appears to derive its revenue through placement of both advertising related content links to partner Web sites within the product.

The terms of service (provided only during installation) discloses that ThirdVoice may release "certain Member Information or any other information about Member's use of the Service" in aggregate, but will exclude the "Member's name, mailing address, email address, and account". The terms of service is a legal document and carries with it legal terminology. Representations of the privacy policy and usage of member information are vague, and users are directed to read the full privacy policy on the Web site.

The privacy policy discloses that non-PII will be collected to determine popular sites, advertisement impressions, and clickthroughs.

ThirdVoice writes that they do not "capture any information that members may submit to the underlying Web site, such as passwords or credit card numbers." This statement is incorrect, since passwords, credit card numbers, and other information could be contained in the query strings that they do capture. Perhaps they mean to write that they will not act upon this information.

ThirdVoice does not provide a means for notifying users of changes to the privacy statement, despite the fact that they collect email addresses from users at the time of registration. The terms of service document does specify a mechanism by which new terms will be communicated to users, but the privacy policy seems to be maintained separately.

CHOICE

The user can disable tracking statistics that ThirdVoice collects by turning off the software.

Email notifications "from ThirdVoice or its selected partners" are offered on an opt-in basis from the "My Profile" page on their Web site.

Uninstalling the software leaves user-specific data files and registry settings for ThirdVoice intact. No mechanism is apparent to delete information gathered about a user in the remote ThirdVoice database.

ACCESS

A profile page on ThirdVoice's site allows for editing a user's name, gender, year of birth, country, and zip code after they have been entered.

There is no apparent access to other information that may be gathered by ThirdVoice about a user's account and usage.

SECURITY

We did not observe the use of encryption between the user's computer and the ThirdVoice server.

MISCELLANEOUS

The practice of changing the appearance of copyrighted Web pages for business purposes, even if only slightly, raises some interesting legal questions.

RECOMMENDATIONS

1. Disclose why this product is being made available free of charge.
2. Improve vague statements about usage of member information in the terms of service.
3. Fix incorrect statement about not capturing information submitted to underlying Web sites.
4. Provide a mechanism to alert users when the privacy policy changes.
5. Disclose that the software is in communication with the ThirdVoice server as long as the user is active on the Web.
6. Remove query strings before transmitting URLs to the ThirdVoice server.
7. Do not transmit information about HTTPS, FILE, and RES URLs to the ThirdVoice server.
8. Provide a mechanism permitting users to terminate their subscription to the service and delete information obtained about them.
9. Provide a mechanism giving users access to the data gathered about them by the service.
10. Use SSL or something comparable to protect communications with the ThirdVoice server.

PRODUCT DESCRIPTION

"Zack is an online service that delivers useful e-commerce features to you directly on a given product page through an interface called the Zack Bar™. In addition to the bar, we also create a personalized myPage™ for you automatically that is available at www.zack.com whenever you are logged on. On your myPage, you can access additional Zack features and obtain one-click access to retail sites supported by Zack."

Version tested: July 21, 2000: http://www.zack.com
Privacy policy obtained August 21, 2000: http://www.zack.com/about/privacy.html
Terms of use: http://www.zack.com/about/terms.html

DATA FLOW

During installation, the user must provide a username and password; the email address is optional. The user logs on to the ZackBar with the username and password in order to start a session.

The ZackBar installation includes an ActiveX control that changes the user's Internet Explorer proxy settings to point to a proxy server at Zack. This server analyzes HTML traffic, looking for keywords that indicate shopping activity. On allied sites, the Zack proxy inserts HTML content showing alternate shopping opportunities and referring back to the Zack service.

Specifically, the ZackBar redirects most HTTP traffic through the remote Zack server, giving the server complete access to the user's full clickstream including query strings, all cookies set and/or transmitted to web sites, all HTTP authentication exchanges (including login/password pairs), and all user data submitted in HTTP forms. The user's Zack ID is not included in the ordinary proxy stream. HTTPS and other protocols are neither proxied nor monitored by the Zack service, nor are transactions concerning files unlikely to contain HTML (such as .doc and .gif files). However, every HTTP transaction containing a query string is specifically selected for analysis.

Since the Internet Explorer proxy settings are machine-wide, HTTP traffic originating from other applications is also routed through the Zack proxy.

We observed Zack inserting a Web bug on one commercial Web page that transmitted the user's Zack ID back to the Zack service. The role of this Web bug was not clear to us, but it may be used to alert Zack which subscriber's purchase is imminent.

NOTICE

The product is presented as free to the user. It appears to steer users towards merchants who are affiliated with Zack.

Since Zack's information practices are presented as optional reading, users may not believe they have agreed to them.

Zack writes "we are committed to staying on the cutting edge of privacy protection," and "it is our intention to collect information that is related only to providing our feature set." This commitment and intention is hard to reconcile with the policy of collecting almost all of a user's Web traffic, given that competing products provide comparable features without such extreme

measures. The privacy policy does make it clear that Zack inspects all Web traffic, but this is not disclosed outside of the privacy policy, such as during installation time.

Zack acknowledges that personal information may be embedded in URL query strings but grants themselves carte blanche, writing "We are not responsible for protecting any personally identifying information transmitted to us in this way." In the next paragraph they write that they will not provide PII or aggregate data to third parties.

The policy states that "We do not capture or store any personal information that is transferred to our offices by our products or services, unless it is included in the URL itself." This cannot be correct, since the proxy captures all HTTP Web traffic by design and cannot immediately distinguish between data to store and data to discard.

The disclosure that the registration cookie "is a small data file on your hard drive that contains your IP address and browser type" understates the cookie's role. The cookie appears to contain a GUID. In addition, "IP address" is undefined jargon.

Zack does a good job describing how they will use email to notify users of privacy policy changes. However, the terms of use state that Zack may modify its services without notice.

In an interesting perspective shift, Zack discloses that they "cannot guarantee or warrant the security of the information you transmit to us". Does Zack equate traffic captured due to use of the ZackBar with "information you [the user] transmit to us"? Most users are unlikely to consider traffic monitored by the Zack service as information "transmitted to" Zack.

In the FAQ, Zack states "there is nothing to download. You simply enter in your name and password and start browsing". However, installing Zack does require downloading the ActiveX control that changes the user's proxy settings.

CHOICE

Users do not have to provide an email address during registration; if they do, the address is only used to communicate "changes that impact the service (changes to the privacy policy, etc.)". Other uses of the email addresses are on an opt-in basis.

Users may terminate their account by sending an email or surface mail request, although it is unclear to what extent this purges data associated with the user.

Zack is not listed under the standard "Add/Remove Programs" dialog. We could only get rid of it by hunting down and removing the appropriate ActiveX DLL. This is not an acceptable uninstallation method for ordinary users, and it severely limits the user's choice in participating.

ACCESS

The Zack software keeps track of products viewed recently and allows the user to delete items from this list. No other mechanism is apparent for viewing data gathered about a user.

SECURITY

The ZackBar sends Web traffic through another site without encrypting it, providing an opportunity for eavesdroppers along the new path in addition to new failure modes. The ActiveX

control that changes the proxy settings is scriptable, so adversaries could use it to interfere with the user's browsing. Fortunately, the control does not directly allow an adversary to specify an alternate proxy server.

MISCELLANEOUS

Logging out of the service via Zack's Web site turns off the proxy settings until the user logs in again, but we were unable to find this documented anywhere.

The practice of introducing foreign content and changing the appearance of copyrighted Web pages for business purposes raises some interesting legal questions.

RECOMMENDATIONS

1. Make the privacy disclosure more prominent.
2. Disclose why this product is being made available free of charge.
3. Fix incorrect statements about not capturing any personal information, understating the cookie's role, and there being "nothing to download" in order to use Zack.
4. Define jargon terms.
5. Reconcile conflicting statements in the privacy policy and terms of use about notifying users of privacy policy updates.
6. Ensure that information previously gathered is destroyed when a user requests account termination.
7. Provide an entry in the standard "Add/Remove Programs" dialog to remove Zack from a computer.
8. The policy of specifically selecting query strings for analysis is inappropriate in light of Zack's disclaimer of responsibility towards data gathered using this technique. Adopt a privacy-friendly policy regarding query strings.
9. Changing a user's proxy settings is an extremely invasive act that calls for correspondingly extreme explanations and warnings. The ActiveX control that performs the change should not be scriptable and there should be a convenient method to completely remove it from a user's system. Basically, the entire install/enable/disable/uninstall design should be revisited.
10. We strongly recommend the adoption of a less invasive data flow architecture for subsequent releases of the ZackBar software.

VII. RESOURCES

ABOUT THE AUTHORS

This research is a project of the University of Denver Privacy Center. Collaborating with the Privacy Foundation, the Privacy Center addresses electronic privacy topics from technical, legal, business, and social perspectives.

David M. Martin Jr. (dm@cs.du.edu) is an assistant professor in the department of Mathematics and Computer Science at the University of Denver.
Richard M. Smith (rms@privacyfoundation.org) is the Chief Technology Officer of the Privacy Foundation.
Michael Brittain (mbrittai@cs.du.edu) is an M.A. student in Digital Media Studies at the University of Denver.
Ivan Fetch (ifetch@cs.du.edu) is a B.S. student in Computer Science at the University of Denver.
Hailin Wu (hwu@cs.du.edu) is a Ph.D. student in Computer Science at the University of Denver.

ACKNOWLEDGMENTS

We gratefully acknowledge the contributions of Reneé Albersheim, John Boak, Justin Rickard, Robert Roberts, Andrew Schulman, John Soma, and Sandy Wright in the preparation of this report.

GLOSSARY

Clickstream. A series of URLs over time corresponding to a user's Web browser use (normally generated by clicking on a series of hyperlinks).

Cookie. A mechanism by which a particular Web browser (and usually a user) can be recognized by a Web site. See §II.A.4.

Ethernet address. A unique number built into an Ethernet network adapter. See §II.A.6.

First party. The computer issuing a Web request, i.e., an end-user's computer. See §II.A.5.

GUID. A Globally Unique Identifier as assigned by *guidgen* or a comparable tool. See §II.A.6.

HTML. Hypertext Markup Language, the primary language used to describe Web page layout.

HTTP. Hypertext Transfer Protocol, the primary protocol used to transmit Web pages over the Internet.

HTTPS. Secure HTTP. URLs beginning with the string "https://" refer to Web resources that are obtained by using *SSL/TLS* instead of an ordinary unsecured connection.

IP address. Internet Protocol address. A number, based on the location of your connection to the Internet, that identifies your computer while it is online.

MAC address. Media Access Control address. Typically used as a synonym for *Ethernet address*. Not to be confused with a Message Authentication Code (MAC).

PII / Non-PII. Personally identifiable information / non-personally identifiable information. See §II.A.1 and §II.A.2.

Query string. Part of a URL, usually following the '?" character, that is given to a program on a Web server for further processing. Query strings can contain sensitive information. See §III.D.

Second party. The computer immediately responding to a Web request, i.e., the Web server requested by a first party. See §II.A.5.

SSL. Secure Socket Layer. A protocol for transmitting information between two computers while preserving secrecy, authenticity, and integrity against attempted eavesdropping or tampering. See also TLS below.

Third party. A computer introduced into a Web transaction by a second party and usually without the first party's explicit consent. See §II.A.5.

TLS. Transport Layer Security. Essentially a new (and more accurate) name for SSL.

UID. Unique Identifier. Any identifier assigned in such a way that it uniquely identifies a computer or user. See §II.A.6.

URL. Uniform Resource Locator. A string, such as "http://www.privacyfoundation.org/", that identifies a Web page or other Web resource. URLs can also contain query strings. See §III.D.

REFERENCES

[1] E. Mendelson, *30 Ways to Browse Better*, PC Magazine 19:18, pp. 180-203.

[2] *Privacy Online: A Report to Congress*, U.S. Federal Trade Commission, June 1998. http://www.ftc.gov/reports/privacy3/index.htm

[3] *Records, Computers and the Rights of Citizens*, United States Department of Health, Education, and Welfare, 1973.

[4] Report of the U.S. Federal Trade Commission Advisory Committee on Online Access and Security, May 15, 2000

[5] Electronic Communications Protection Act. Title 18, U.S. Code §2510 et seq

[6] The Judnick case against DoubleClick, as well as the other California state cases, have been consolidated into one state action in California State Court. Thirteen federal cases have been consolidated under the judicial panel on multidistrict litigation in the Southern District of New York before Judge Buchwald.

[7] Federal Trade Commission Act. Title 15, U.S. Code §41 et seq

[8] *A Brief Overview of the Federal Trade Commission's Investigative and Law Enforcement Authority*, http://www.ftc.gov/ogc/brfovrvw.htm

[9] Children's Online Privacy Protection Act (COPPA). Title 15, U.S. Code §6501 et seq

[10] August 13, 1998 FTC action against GeoCities: http://www.ftc.gov/os/1998/9808/index.htm

[11] Computer Fraud and Abuse Act. Title 18, U.S. Code §1030

[12] Spyware Control and Privacy Protection Act. 106th Congress Senate bill 3180

[13] Richard M. Smith, *The RealJukeBox Monitoring System*, http://users.rcn.com/rms2000/privacy/realjb.htm

[14] Richard M. Smith, *The Comet Cursor*, http://users.rcn.com/rms2000/privacy/comet.htm

[15] Richard M. Smith, *Alexa and zBubbles*, http://users.rcn.com/rms2000/privacy/alexa.htm

[16] SST Incorporated, http://www.sstinc.com/

[17] July 21,2000 FTC action against ToySmart.com: http://www.ftc.gov/os/index.htm.

[18] TRUSTe privacy seal program: http://www.truste.org/

[19] BBBOnLine privacy seal program: http://www.bbbonline.org/

[20] Platform for Privacy Preferences Project (P3P): http://www.w3c.org/P3P

[21] The ACM Code of Ethics: http://www.acm.org/constitution/code.html

[22] The :CueCat Bar Code Reader, http://www.privacyfoundation.org/advisories/advCueCat1.html