User Perceptions of Privacy and Security on the Web

Scott Flinn and Joanna Lumsden
National Research Council Canada
{Scott.Flinn|Jo.Lumsden}@nrc.gc.ca

Abstract

This paper describes an online survey that was conducted to explore typical Internet users' awareness and knowledge of specific technologies that relate to their security and privacy when using a Web browser to access the Internet. The survey was conducted using an anonymous, online questionnaire. Over a four month period, 237 individuals completed the questionnaire. Respondents were predominately Canadian, with substantial numbers from the United Kingdom and the United States. Important findings include evidence that users have tried to educate themselves regarding their online security and privacy, but with limited success; different interpretations of the term "secure Web site" can lead to very different levels of trust in a site; respondents strongly expressed their skepticism about privacy policies, but nevertheless believe that sites can be trusted to respect their stated policies; and users may confuse browser cookies with other types of data stored locally by browsers, leading to inappropriate conclusions about the risks they present.

Keywords: Human factors, privacy, risk, risk management, security, transparency, trust, usability, Web, Web browser, WWW.

1. Introduction

In the early days of the Internet, a large proportion of its users had considerable technical experience and a deep understanding of its operation. This is no longer true. Today, many Internet users exchange e-mail, view pages and consume services from the World Wide Web, and use it for any of several other common activities. Engaging in these activities can pose a number of risks to an individual's online privacy and security. Without a deep technical understanding to guide them, some risks may be elevated for these users.

The technically skilled people who have designed and built the software used for these activities – e-mail clients, Web browsers, etc. – have equipped the software with various tools that individuals might use to remain informed of the risks they face and to manage various aspects of their privacy and security as they utilize the Internet. However, anecdotal evidence suggests that typical users are unaware of many risks, or misunderstand them, and that they are ill equipped to use available tools to manage those risks.

This paper describes a preliminary study intended to explore some of the assumptions underlying the design of specific privacy and security tools available in contemporary Web browsers. Our objective was, in part, to empirically test common assumptions that are based primarily on anecdotal evidence. Participants in our study responded to a questionnaire designed to highlight the extent of their awareness, knowledge, and range of perspectives concerning the technology.

The questionnaire focused on four specific areas relating to the use of Web browsers: secure Web sites; browser cookies; Web site privacy policies; and trust marks. In each category, it sought to elicit responses that would reveal what users understand about the risks they face when using a Web browser to access the Internet in a variety of situations; how aware they are of the tools at their disposal for managing security and privacy risks in those situations; and how prepared they are to use the tools in terms of both willingness and ability.

The survey was largely qualitative and was not designed to produce data for statistical (other than descriptive) analysis. Our objective was not to precisely characterize a particular user population, but rather to explore some of the misunderstandings and misinterpretations that can lead to undesirable but avoidable outcomes. Ideally, we also hoped to gain some insight into how those inaccurate perceptions arose.

Similarly, the questionnaire was not intended to assess the usability of particular technologies, but rather to determine which technologies are actually used by our respondents.

The study consisted of a single questionnaire that was advertised publicly on the Internet and administered anonymously online. We received 237 responses over a period of four and a half months from 7 June 2004 to 27 September 2004. Section 5 presents additional summary statistics for the questionnaire along with our detailed findings.

Our survey was essentially designed to answer the following questions:

1. Are typical Internet users fully aware of the risks they face online, and do they understand these risks in sufficient detail and with sufficient clarity to support accurate risk management decisions?
2. Are typical Internet users aware of the common Web

browser tools available to assist with this decision making process, and do they use them when appropriate to make accurate and informed decisions?

Based on anecdotal observations, our expectation was that the survey would highlight low levels of awareness and understanding of online security issues amongst typical Internet users, and a subsequent avoidance or inappropriate use of available tools to manage online security risks.

2. Related work

A growing body of evidence clearly demonstrates that Internet users tend to base their judgments of the trustworthiness of Web sites on characteristics such as navigation and fulfillment (*e.g.*, [2]) that are not causally linked with the actual trustworthiness of the site. They also tend to be too relaxed in terms of critical thinking, preferring instead to focus on prominent features without considering their importance very deeply [6].

On the other hand, they are clearly concerned about their own privacy and safety, and the privacy and safety of the information that describes them and of the technology they use [4][8].

The general problem, then, is that users are concerned about the risks they might face online but do not currently make good risk management decisions. Many technologies have been proposed to address the problem, but often they seem to compound it. For example, it is straightforward in principle to encrypt and digitally sign e-mail, but even experienced computer users have difficulty performing the simplest of these tasks [12]. This result has been widely cited, and the situation has not improved significantly in the six years since it was published. Less empirical evidence is available with respect to the use of Web browsers to conduct business with Web sites through secure connections, but there is clear cause for concern [5][9][11].

A reasonable conjecture is that the technologically minded people who have designed and built our networks and applications have produced an environment that can only be effectively understood and controlled by those who think in similar ways. It is unreasonable to expect users to think like Computer Scientists and Engineers when they go online, and not surprising when they don't. For example, there is a stark contrast between the human terms in which a sociologist describes trust online [7] and the algorithmic trust models on which the technology is built (*e.g.*, [1]).

It is relatively straightforward to document the symptoms of the problem, but more of a challenge to understand the underlying causes. The study we present here is intended as a preliminary step toward that understanding.

As indicated above, the purpose of the study was to look specifically at how people perceive certain limited classes of risks, and how they employ specific software tools to help them manage those risks. It was not intended to assess the usability of the tools, nor to identify broader or more general issues. It builds on previous research rather than repeating or seeking to reproduce it. For example, work has been done to identify the range of concerns expressed by typical Internet users [4][8]. Many of the questionnaire's response options were guided by these findings to help ensure that closed-ended questions offered a meaningful set of choices. The same research has also revealed differences between the perceptions of different demographic groups; these findings guided the choice of specific demographic questions for our questionnaire. Research has been done to explore how users relate to and use specific software features. For example, Millett et al. have performed a retrospective analysis of tools for managing browser cookies [11]. Friedman et al. have performed user studies to determine how typical Internet users assess whether a given connection to a Web site is secure [9]. The findings from these studies guided the development of questions in our study relating to cookies and secure Web sites.

Very little work has been done, however, to assess the effectiveness of existing technology. Cranor et al. have performed such analyses in the domain of Web site privacy policies [3], but focused primarily on a piece of software that is not included with any browser by default and which is not widely deployed. Recently, using a survey similar to the one described here, Garfinkel et al. [10] explored the perceptions and opinions of online merchants regarding digitally signed e-mail. Although some of their general findings *are* relevant in this context, their focus was on a distinctly different piece of end-user security technology; the need, therefore, remains to explore browser related technologies more fully.

3. Survey design

The questionnaire began with a set of demographic questions about age, education, and country of residence. These were followed by a few very general questions to determine which browser and operating system respondents were accustomed to using, and the frequency with which they use them.

The remainder of the questionnaire repeated the same pattern of five questions for each of the four technologies of interest (secure sites, cookies, privacy policies and trust marks). The first question in each section asked whether the respondent had any previous knowledge of the technology; only when respondents indicated that they had previously heard of it were they required to complete the rest of the questions in the associated section.

The second question in each section asked respondents to describe in a few brief sentences what they understood about the technology. The third probed their beliefs about the technology by listing a number of statements pertaining to the technology and asking them to indicate the degree to which they agreed or disagreed with each statement; the statements and the five-point Likert scale response options for each were collectively presented using a matrix-style format.

The fourth question assessed respondents' familiarity with the technology in question, and the fifth explored the degree to which respondents' feelings of security and privacy depend on the technology.

There were 69 questions in total, including the four open text questions. Most respondents answered the majority of the questions, taking about 20 minutes on average to complete.

3.1. Recruiting

We invited people to respond to the questionnaire by circulating a standard invitation to participate via e-mail. Our target audience was Internet users, and, ideally, we hoped to recruit a representative sample. Through friends, family, and colleagues, we circulated the invitation on mailing lists and in discussion fora whose purposes were not related to information and computing technology. We also invited geographically distant friends and relatives to participate, and to relay the invitation to their own contacts.

Section 5 summarizes the outcome of this recruiting effort, including a characterization of the group that ultimately responded.

3.2. Implementation and mechanics

The questionnaire was implemented as a set of servlets and Java Server Pages (JSPs) running in the Apache Tomcat servlet container on our server. The recruiting message directed potential respondents to a consent form. Those who declined consent were instructed simply to proceed no further. The remainder indicated consent by clicking a clearly labeled form submission button. When respondents indicated their consent, they were assigned a user number, counting sequentially from one.

We relied on Tomcat's HTTP session tracking, supplemented by URL rewriting to support those users whose browsers blocked Tomcat's session tracking cookies. Sessions were set to expire in 60 minutes. Requests for questionnaire pages that were not associated with an active session were redirected to the consent page. During the questionnaire period, we registered 23 session expiries. Of these, only one individual persevered to give their consent a second time and complete the questionnaire. We did not record any instances of someone attempting to fetch questionnaire pages without first giving consent.

Although we did not expect any attempt by a malicious respondent to skew the results by submitting multiple responses, we nevertheless made an effort to detect multiple responses from the same source to assure the quality of the results. We accepted all responses and labeled each response set with a value derived from the browser source IP address and user-agent string. This allowed us to detect likely cases of submission from the same browser (though it would not thwart a determined adversary who varied the user-agent string or spoofed the source IP). Several such duplicates were in fact detected, though no more than four responses came from the same IP/user-agent combination. We inspected each of these responses individually and did not detect clear evidence of multiple responses from a single individual. In most cases, the responses from the same IP/user-agent combination were received on different days. Likely explanations include IP addresses being re-used by respondents who share the same ISP, responses from public access terminals, or responses from individuals within an organization with a standard browser deployment and a shared HTTP proxy.

To help preserve anonymity, our server was configured to not log requests, thereby preventing source IP addresses from being recorded. The browser "fingerprint" described in the previous paragraph was computed by concatenating a string representation of the source IP address with the user-agent string as received from the agent, and computing a SHA1 digest of the resulting string. Only the digest was stored for comparison. Given the wide variety of user-agent strings in use, even for what is ostensibly the same browser (predominately Microsoft's Internet Explorer version 6), and combined with the relatively large space of source IP addresses, this technique makes it computationally expensive to recover IP addresses for respondents, while still allowing a comparison based on source IP address.

As respondents progressed through the sequence of questionnaire pages, their intermediate responses were recorded in their HTTP session record. If a session expired or was abandoned, these responses were lost. Upon reaching the end of the questionnaire, respondents were instructed to click a button labeled "Submit Responses" to have their answers recorded, or to simply close their browser window to have their answers discarded. Upon clicking the submit button, their data was copied from the transient session record into the database.

4. Did design affect results?

In many cases it was difficult to formulate questions with enough precision to be useful, but without revealing too much of the expected answer in the question itself. Early drafts of the questionnaire used precise terminology and its questions were very specific. Reviewers of these drafts quickly pointed out that, in doing so, the questions revealed sufficient information that they were likely to influence respondents' answers. The broad range of technical expertise that we expected to encounter compounded the difficulty. The final questionnaire consistently errs in the direction of imprecision when necessary to avoid biasing the responses.

Nevertheless, there is evidence that respondents still based their responses on information they learned or deduced from previous questions. For example, several respondents reported that they were unaware that they could click on trust marks to authenticate them, yet later reported that they are more likely to trust a site if they authenticate the trust mark it displays by clicking on it.

This section discusses these and other ways in which design decisions influenced the responses we received.

4.1. Secure Web sites

One of the most interesting findings of the survey concerns how people interpret the phrase "secure Web site". In technical circles, the term commonly refers to the use of SSL/TLS technology for encrypting and authenticating HTTP connections. Indeed, the majority of respondents appear to have used this interpretation in answering related questions. That said, many people interpreted the term to mean that the site itself was secure in some unspecified way. This is not surprising given that the term refers to a *site*, not a *transport channel*. Users who chose this interpretation demonstrated a significantly different understanding. The issue is discussed further in Section 5.

As a design issue, the first technical question asked

respondents how familiar they were with the term "secure Web site". Early drafts of the question spelled out what we meant by secure site before asking about familiarity with the term. In the final version, interpretation of the term was largely up to the respondents, giving them the flexibility to tell us their understanding of the term and for us to elicit a range of possible interpretations.

4.2. Browser cookies

One question asked respondents whether they believe cookies help improve their browsing experience. A large majority of respondents agreed, as we expected. But one respondent indicated that he disagreed with the statement because he believes that it should be technically possible to make a site just as usable, friendly and customized without using cookies. The same may have been true for other respondents who disagreed but did not explain their reasoning in the open text portion of the question.

One question asked respondents to indicate their level of agreement with the following statement:

> Cookies can reveal to a Web site the names of other Web sites I have visited.

This is a good example of a question that would be difficult to make more precise without revealing too much of its motive. Strictly speaking, cookies themselves do not reveal names of other sites; it is the *referer* header in the HTTP request that does so. Cookies are implicated, however, because they potentially allow individual requests to be linked to build a list of names from the *referer* values of other sites the user has visited.

We did not ask respondents about the distinction between persistent cookies and session cookies. At least one respondent observed that the correct answers to some of the questions depend on which type of cookie is being considered.

4.3. Privacy policies

One question asked respondents to indicate their level of agreement with the following statement:

> A privacy policy helps protect sensitive information after it has been collected by a Web site.

Strictly speaking, the privacy policy itself cannot protect information after it has been collected, but it may provide assurance that the operators of the site will take other measures to protect the information. Answers to this question will depend on which interpretation respondents chose.

4.4. The meaning of privacy

One question asked respondents to indicate their level of agreement with the statement, "Cookies invade my privacy". Some responses were surprising. For example, several respondents strongly disagreed that cookies invade their privacy, but agreed with all of the statements positing that cookies reveal personal information to Web sites and allow browsing behavior to be tracked.

It is clear that the interpretation of the term *privacy* will influence many of the responses to questions about cookies and privacy policies, but we did not ask questions explicitly designed to reveal respondents' interpretation of privacy.

The questions relating to cookies and privacy policies were not precise about certain distinctions out of concern that more precise questions would be too technical or too revealing. In particular, our questions relating to privacy policies presumed a P3P-like model, but did not name P3P explicitly. The distinction between compact policies and full policies was therefore omitted entirely.

With the deployment of P3P capabilities in contemporary browsers, cookie management has become implicated with the enforcement of privacy preferences. This connection is not reflected in any of the questions relating to browser cookies or privacy policies, as we felt such questions would be incomprehensible to a lay audience. In retrospect, it seems likely that many of our respondents would have been knowledgeable about the distinction. Had we asked about it, we may have learned more about the range of perspectives that might ultimately filter down to less technical users.

4.5. Self-selecting respondents

Although we had some control over how we seeded our recruiting message, we had little insight into how it propagated, and our respondents were all self-selected (as is the norm for most online questionnaires).

It is not our intention, however, to use our data to characterize the awareness, knowledge or beliefs of a more general population, but rather to reveal a broad range of real beliefs. Knowledge of what interpretations (or misinterpretations) are possible will be valuable in designing privacy and security tools that are effective across the *spectrum* of potential Internet users.

The open text responses were critical to the success of the study because they frequently provided explanations for why respondents answered as they did. Without the text responses, the questionnaire data would have been much less revealing.

In many cases, it would have helped to know whether the respondent was a domain expert in the area (as many apparently were, based on their text responses). We did ask questions about education level and experience with computers. In retrospect, it would have been useful to ask at least one additional question to assess technical expertise in areas relevant to the questionnaire.

4.6. Browser identification

For each technology, we asked respondents to indicate their awareness of specific browser features. We can assess the accuracy of these responses based on which browser they have in mind when answering the question. In most cases, the presence of a feature depends on the browser version as well as its vendor. For this purpose, we preferred to use the user-agent strings sent by browsers rather than relying on respondents to accurately report details of browser version. However, since there was a possibility that respondents would fill out the questionnaire using a browser that is not the one with which they are most familiar, we asked respondents to indicate whether they were currently using the same browser that they most often use for Internet activity. For

affirmative responses, we have assumed that the user-agent string then accurately identifies the browser with respect to which subsequent questions were answered. For negative responses, we asked respondents to select the name of the browser they normally use, without reference to a specific version. When assessing the accuracy of responses related to specific browser features, we used only the first group of responses.

4.7. Other limitations

One respondent neatly summarized some of the methodological limitations in saying, "*The correct answer to many of these questions relies on your browser settings*". In our analysis we have looked for broad trends and a range of different perspectives, and have avoided making precise quantitative statements other than simple proportions.

5. Results and analysis

The survey ran from 7 June 2004 to 27 September 2004. During that time, about 470 visitors viewed our consent page. Of those, 356 gave their consent to participate and 237 persevered to the end, submitting answers to the questionnaire. One respondent skipped every question, completing the questionnaire in 17 seconds. This response was discarded (treated as if consent was given but no responses submitted), leaving 236 complete responses. The average time taken to complete the questionnaire was about 20 minutes. We logged a total of 109 hours of interaction with our questionnaire (more than 4.5 days). Those who completed the questionnaire account for more than 77 of those hours (more than three days of effort). They registered over 15,000 answers and opinions through radio button responses and contributed over 22,000 words through their text responses. Needless to say, we are very grateful to all those who participated.

Tables 1-5 provide some summary statistics describing participation in the survey. A technical report (available from the authors) containing an expanded version of this paper provides a more complete description of our sample, as well as the full text of the questionnaire and a detailed quantitative summary of responses.

Table 1: Country of residence (as reported)

Country	Responses	% of Total
Canada	172	72.6
United Kingdom	31	13.1
United States	17	7.2
Spain	3	1.3
Israel, Italy	2	0.8
Others	1	0.4

Table 2: Responses by age group

Age Group	Responses	% of Total
18 to 20	6	2.5
21 to 30	78	33.1
31 to 40	78	33.1
41 to 50	41	17.4
51 to 60	27	11.4
61 to 70	4	1.7
71 or older	1	0.4

Table 3: Responses by gender

Gender	Responses	% of Total
Female	78	33.1
Male	157	66.5
Unspecified	1	0.4

Table 4: Responses by operating system

Operating System	Responses	% of Total
Windows (all versions)	190	80.5
Linux Intel	20	8.5
Mac OS (all versions)	17	7.2
Unknown	9	3.8

Table 5: Responses by browser type

Browser	Responses	% of Total
MSIE	153	64.8
Mozilla-derived	58	24.6
Safari	10	4.2
Other	15	6.4

Among these demographic statistics, Table 5 (browser types, as reported in the *user-agent* request header) contains a hint as to how representative our sample was. Various sources report the market share of Microsoft's Internet Explorer (MSIE) at well above 90%. For example, for June 2004 – the period during which we recorded most of our responses – The Counter (http://www.thecounter.com/) reports a 94% market share for MSIE, with Mozilla derived browsers at less than 4% and Apple's Safari at just 1%. At the time, Google's Zeitgeist (http://www.google.com/press/zeitgeist.html) reported similar figures for the same period. The profile of our respondents much more closely resembles the readership of the Safalra site (http://www.safalra.com/website/browsermarket/index.html), which features information about hypertext, programming languages, and general science topics. On this basis, it seems reasonable to speculate (but not to conclude) that our respondents were more representative of a technically sophisticated audience than of the wider population of Internet users.

5.1. Analysis

To support the data analysis, we created a servlet capable of extracting subsets of responses based on demographic and other criteria and displaying summary statistics for each question. The format of the summary page mirrors the original questionnaire, displaying each question and all of the response options. For each response option, it displays the percentage of respondents in the subset who selected that option, and color-codes the option such that higher percentages result in darker colors. This proved to be an effective way to quickly discern the pattern of responses for any given subset. Subset criteria could be quickly specified using a simple hand-crafted HTML form.

For subsets containing a single response, the response to each question was immediately evident as only one option was darkly colored in each case. This served as the basis for examining individual responses. The consistency of presentation between responses was useful in recognizing patterns of responses. Open text responses were displayed in full for subsets having only one response; for larger subsets, only the numbers of open text responses given for the question were reported.

Following the data collection period, we began the analysis by reading through the responses individually and making notes about potentially interesting findings and apparent trends. Based on this preliminary analysis, we partitioned the responses in a number of ways and examined the response patterns for each group. Specifically, we examined three partition criteria.

Interpretation of "secure site". We asked respondents to tell us, in their own words, what they understood the term "secure Web site" to mean. It was immediately evident from the responses that two quite different interpretations of the term were common. As noted earlier, some respondents thought that it referred specifically to the securing of transport connections using SSL/TLS (the "secure connection" interpretation), while others thought that it referred more generally to the security of the site itself, its hosts, servers and databases (the "secure site" interpretation). In many cases, the text responses were unambiguous in stating one interpretation or the other. We used these text responses to manually divide the 236 responses into three categories:

- those clearly stating the "secure connection" interpretation (96 responses);
- those clearly stating the more general "secure site" interpretation (53 responses); and
- those that did not provide a text response, or whose response left any room for uncertainty (87 responses).

The pattern of responses to the opinion questions regarding secure sites was clearly different between the two groups, as discussed below.

Browser and operating system. The overwhelming majority of Internet users use Microsoft Internet Explorer running on some version of Microsoft Windows, while more security-conscious users often make other choices. To explore possible trends, we created the following partitions:

- respondents using Windows, any version, any browser (178) vs. those not using Windows (39);
- respondents using Internet Explorer (143) vs. those using another browser (74); and
- respondents using Internet Explorer on Windows, any version (142) vs. those using a different operating system *or* a different browser (75).

The second and third divisions were nearly identical because of the strong coupling between Internet Explorer and Windows (we had only one respondent using IE on a Macintosh). Note that complementary sets add to 217 responses, not 236. As noted earlier (Section 4.6), we used the user-agent string for this partitioning and only included those respondents who indicated that they were answering the questionnaire using the same browser they normally use for Internet activity. These divisions did not reveal any noteworthy patterns.

Confidence in responses. Several questions in each section asked respondents to indicate how confident they felt about their knowledge of the technology, or about adequately managing their risks. We used these responses to partition respondents into those who were confident and those who were not with respect to each technology. Because respondents often expressed confidence in one area but not in another, we did not feel it would be meaningful to average confidence responses across technology categories. Consequently, we ended up with a relatively large number of groups, none of which revealed strong response patterns.

5.2. Findings

Security practitioners, especially those on the front lines, frequently lament that users are uninformed and unmotivated regarding security issues, and call for renewed efforts to educate users. While we too are convinced that education is a critical component, our survey provides some evidence that the benefits of modest education may not be as dramatic as we would hope. The problem may not be so much that people are not interested in learning, but that it is a difficult subject. We provide many examples in this section of highly educated users who have clearly made an effort to understand the technologies they use, but who nevertheless retain serious misconceptions.

For example, one respondent reported – through an open text question – that he does not understand what a secure Web site is. He has read information about site security but has not retained it:

> *"My only knowledge of secure web sites is that they store sensitive information on a seperate [sic] secure server however I'm not really sure what that means or how it benefits me. I have read the security information provided on a few secure sights [sic] but I have not retained the information, possibly due to not fully understanding it."*

The evidence is most pronounced for matters relating to browser cookies (see Section 5.2.2 below).

Because we cannot say how representative our sample is, we cannot draw strong conclusions about Internet users in general. However, we believe the evidence we present here is sufficient to raise a question about the efficacy and potential of widespread education about personal privacy and security.

As expected, correlation between education level and

technical knowledge was weak. Most of our respondents were highly educated, with 82% having completed some kind of post-secondary training, and 41% possessing an advanced university or professional degree. Many respondents in these groups had only superficial understanding of the technologies they used.

5.2.1. Secure sites

Only one respondent (0.4%) admitted to never having heard the term "secure Web site". Another 12% had heard of the term, but had little or no idea what it means to be secure. 88% claimed to have at least some knowledge of secure sites.

Transport vs. storage. As noted earlier, respondents interpreted the term "secure Web site" in one of two ways. Some assumed it referred specifically to HTTP over SSL/TLS, while others assumed it referred more broadly to the security of the entire site. The interpretation had a significant impact on subsequent opinion questions. For example, we asked respondents to indicate their level of agreement with the statement, "A secure Web site assures me that the site is trustworthy for the purpose of conducting business". Among respondents who clearly used the "secure connection" interpretation, 61% disagreed to some extent, while 18% agreed. (Respondents who chose "strongly disagree" or "disagree" for a given statement are said to disagree *to some extent* with that statement, and similarly for agreement). Among those who clearly used the "secure site" interpretation, in contrast, only 18% disagreed to some extent while 55% agreed. In other words, those with the "secure site" interpretation were much more likely to regard a site as trustworthy.

Many respondents in the latter group reinforced these views in their open text responses. One respondent characterized a secure site as "*A site [where] I can carry out business transactions with confidence*". Another put it this way:

> *"The information given on a secure web is for the recipient only and cannot be shared or stolen. It makes buying on the internet a much safer experience."*

In general, respondents in this group had a greater number of misconceptions about the assurances provided by Web sites labeled as "secure". One way to interpret the responses we received is to postulate that many people do not clearly perceive the distinction between transport and storage. For example, consider these statements:

> *"When a website is secure, other people can't see your credit card numbers, personal info., etc. when ordering things online."*

> *"Information is encrypted to preserve privacy."*

There is no hint here that the respondents are thinking about the distinction. On the other hand, it was widely understood (using either interpretation) that a secure site involves encryption in some way. How then would users who do not consciously consider the distinction between transport and storage interpret the message that a closed lock icon indicates a "secure site"? One possible answer is that they will be confused about what data is encrypted, where, and for how long, and consequently conclude that the lock icon indicates that information they submit will be permanently encrypted at the server. We found substantial evidence to support this hypothesis. For example, one respondent characterized secure sites this way:

> *"The servers are in a secure location, and data is encrypted by very high level (eg256 bit) encryption. Concern: nothing is infallible, and geeks can crack what geeks created."*

Several respondents indicated their belief that personal information submitted to a Web site is selectively encrypted according to its sensitivity. For example, one respondent said:

> *"I'm under the impression that with secure websites, any personal information that I may enter is only accessible to the compnay [sic] that I intend to provide the information to and that things like social security numbers or passwords are encrypted."*

Another believes that "*login id and passwords are encrypted when transmitted*". Some users simply place trust in a higher power:

> *"I think it means that the information I give to the website can't be accessed by anyone else. I hope that's what it means!"*

> *"I understand that information sent between the client and server is encrypted using a very strong encryption method. My major concern is that the data stored on the host computer may be stolen. However, I'm fairly confident that major institutions, such as banks, have this problem licked."*

The problem is compounded because we in the technical community have used the term "secure site" to refer to something very specific (HTTPS transport) that cannot be deduced from the label itself. It is not surprising that many people interpret the term "secure site" to mean a *site* that is secure.

While those respondents who associate security only with the transport connection often demonstrated detailed technical knowledge of the protocols involved, there were many in this group who were clearly less knowledgeable. One respondent knew that the term "secure site" somehow implicates encryption, but was uncertain precisely how:

> *"I think secure Web site use encryption when sending information. But I am not quite sure what encryption really means, and if certain people can still intercept that information and make use of it."*

Friedman et al. have also drawn attention to differing interpretations of Web site security [9]. In their study, they asked subjects to define "secure connection", and divided responses into three categories: *transit*, *encryption*, and *remote site*. They found roughly similar proportions, though a precise comparison is not meaningful because of differences in the sample populations. Their study did not explore a connection between interpretation and other beliefs about secure sites.

Untrustworthy transport. A surprising number of people

disagreed with the following statement:

> I can *always* rely on a secure Web site to protect sensitive information as it is being sent to or from the site.

It is surprising because protecting information in transit is a relatively strong link in the chain and one of the few things for which HTTPS *can* be solidly relied upon. The result does not appear to stem from the interpretation of "secure site". Over 35% of the respondents with the "secure connection" interpretation still disagreed to some extent, though over 52% agreed. Over 35% of the respondents with the "secure site" interpretation disagreed to some extent while 41% agreed – similar proportions. So what were the 35% of respondents in the first group thinking?

The most likely explanation is that many respondents were influenced by the italicized word *always* in the statement. One respondent made the following observation:

> *"... any questions that say "always" are useless; all generalities are false and I must disagree with them on principle."*

Another respondent explained in the open text response that he had to disagree because one cannot be certain of perfect confidentiality:

> *"A secure web site establishes an encrypted channel (usually using SSL), through which HTTP can be sent back and forth without the possibility of a third party easily intercepting and reading the communication. When using a secure site, we can be more confident (though not perfectly confident) that sensitive information is being seen only by the intended recipient."*

Awareness of server authentication. Many respondents, including those who appear to be relatively well informed about the technology, seem unaware that the SSL/TLS protocols underlying HTTPS provide server authentication. We asked respondents to indicate their level of agreement with the following statement:

> A secure Web site assures me that I am communicating with the real site and not an impostor.

37% of all respondents disagreed to some extent, while 37% agreed. Surprisingly, 41% of those who used the "secure connection" interpretation of secure site disagreed, while 37% agreed. We interpret these results to mean that the respondents who disagreed are not aware of the authentication component of secure connections (note that the question does not say *always,* but rather asks whether secure sites give some assurance in this regard). Unfortunately, the open text responses did not reveal an explanation for the trend. One respondent reported the incident some years ago in which VeriSign issued two code signing certificates in Microsoft's name to a fraudulent applicant (complete with a link to a mailing list discussion of the incident). He cited the incident as evidence that a digital certificate does not *guarantee* the identity of the remote party in an SSL/TLS transaction. While this explains one respondent's disagreement, it seems unlikely that it would explain the overall trend.

In the absence of a better explanation, the evidence suggests that respondents were unaware of the benefits (or importance) of server authentication in communicating with secure sites, including many respondents who demonstrated detailed technical knowledge of at least some aspects of the SSL/TLS protocol. There is clearly a need for further investigation.

Awareness of tools. 22% of respondents do not know if their browser can display security details relating to encryption and server authentication. A further 24% believe it does, but have not attempted to view them. Among those respondents who used the "secure connection" interpretation, these percentages are 10% and 16% respectively; for those who used the "secure site" interpretation, the values are 30% and 38%.

This suggests that a significant number of Internet users are unaware of the tools involved, although it is unclear how this result should be interpreted. Is lack of awareness of these tools necessarily a negative thing? Do users who view security details achieve better outcomes? It is interesting to note that in designing its new Safari browser, Apple Computer chose to intentionally hide this information. Although the browser is capable of displaying certificate details in certain rare situations, and it provides warnings when server certificates cannot be validated, it is not generally possible to view the certificate or connection details for a page served through HTTPS.

Reliance on tools. 60% of our respondents reported that they would stop using some sites if they were not secure. On the other hand, 14% reported that site security never influences their trust decisions, while another 25% indicated that they are more comfortable with some sites because they are secure, but would probably still use them even if they were not. Among those respondents who used the "secure connection" interpretation of "secure site", these numbers are 70%, 8% and 22% respectively; among those with the "secure site" interpretation, they are 62%, 9% and 28% respectively.

5.2.2. Cookies

Only three respondents (1.3%) admitted to never having heard the term "browser cookie". Another 9% had heard of the term, but had little or no idea what cookies are. 89% claimed to have at least some knowledge of browser cookies.

We asked respondents to indicate their level of agreement with a number of statements. Among these were a number of distinctly negative statements:

- Cookies invade my privacy.
- Cookies reveal my personal information to Web sites without my knowledge.
- Cookies allow others to track my browsing activities on the Internet.
- Cookies can reveal to a Web site the names of other Web sites I have visited.

Overall, there was widespread agreement with all of these statements. The suggestion that cookies facilitate the tracking of browsing activities was the most broadly supported, with 72% of respondents agreeing to some extent.

Users have tried to educate themselves. The open text

question in this section revealed a wide range of ways to describe cookies. Some demonstrated misconceptions (discussed below), but most had at least a kernel of accuracy. For example, one particular respondent appears to know little about cookies, does not know if his browser can do anything with them, and responded to the opinion questions in ways that are arguably wrong. Yet his open text description of cookies is concise and accurate:

> *"I believe they are files containing personal information that other computers (servers) place on my hard drive to identify my machine, and me, when I access their web sites."*

This trend provides some evidence to contradict the widespread belief that typical Internet users have no idea what cookies are. If our respondents were knowledgeable about only one technology, it was almost always cookies.

Cookies speed up web sites. Many respondents expressed the belief that the primary purpose (or primary benefit) of cookies is to speed up web sites. Many went on to explain that the speed up is obtained because cookies allow login forms (and occasionally other forms, such as payment details) to be bypassed. Some also observed that server-side auto-completion of Web forms saves time as well.

Others appear to have confused cookies with data caching. For example,

> *"A cookie stays on your computer so that when you visit that web page again, it loads pictures faster."*

> *"My understanding of cookies is that my computer stores web sites that are used so when I want to view these sites they can be viewed quicker."*

It is interesting to note that in the user interface for recent versions of Microsoft's Internet Explorer, the button to delete cookies is visually grouped with the button to delete temporary Internet files, which together are introduced with the following statement:

> *Pages you view on the Internet are stored in a special folder for quick viewing later.*

This may explain why some users associate cookies with browser caching, which may in turn lead to inappropriate conclusions about their function and purpose, and therefore about the risks they pose.

Cookies protect data in transit. Some respondents agreed with the following statement:

> Cookies help protect sensitive information as it is being sent to or from a Web site.

This may be indicative of a serious misconception. One respondent went on to explain his reasoning, observing that cookies can obviate the need to (re)transmit sensitive data to a site. It is unknown whether others who agreed with the statement had similar reasons.

Revealing information vs. tracking. We asked respondents to indicate the degree to which they agreed with the following two statements:

> Cookies reveal my personal information to Web sites without my knowledge.

> Cookies allow others to track my browsing activities on the Internet.

In practice, cookies rarely contain personal information. Instead, they generally contain a unique identifier that is linked to information stored at the server that was previously submitted by some other means, such as through an HTML form. For this reason, the first statement is arguably false. On the other hand, one of the greatest privacy concerns about cookies is that they can be used to correlate visits to multiple, often independent web sites with a single user (or browser). Therefore, the second statement is arguably true. We were interested in the number of respondents who would perceive this distinction.

Only 25 respondents (11%) disagreed to some extent with the first statement while agreeing with the second. By including neutral responses, we found that 59 respondents (25%) agreed with the second statement, but not the first. Given that many of our respondents demonstrated a deep technical understanding of the technology, these proportions would tend to suggest that the distinction is not widely recognized.

Other observations. Several respondents noted that a possible concern with cookies is that they can be stolen by other sites (presumably through browser security flaws or packet interception).

One respondent described cookies this way:

> *"Contains information that will allow the website to "recognize" you as a returned user. No personal information is stored on the website server."*

There is too little context to be certain, but this may suggest a misconception that makes cookies seem like a good thing by reducing the need for servers to store personal information. This idea presents an interesting contrast to the belief that it is good for servers to store sensitive information because it avoids the need for repeated transmission.

Confidence. 58% of respondents express confidence that they understand how cookies may be used to track their activities online, while only 17% report a lack of confidence in this regard. On the other hand, only 28% of respondents are confident in their ability to "distinguish between cookies that are beneficial and those that may be harmful", whereas 42% are not.

26% of respondents report having used cookie managers, but without confidence that doing so has helped them. Only 29% are confident that use of cookie managers has helped them.

Awareness. 21% of respondents do not know if their browser allows cookie management of any kind. Another 9% do not know if they have used cookie management features, and a further 12% believe they have not used them. Altogether, that represents 42% of respondents whose awareness is low. 55% of respondents reported having used cookie management features.

Reliance on tools. 27% of our respondents reported that they would stop using some sites if they could not control (or block) cookies exchanged with the site. 72% of respondents reported that they have either never used cookie management features (24%), or would not change their browsing behavior even if cookie management features were not available (48%).

5.2.3. Privacy policies

Thirteen respondents (5%) reported never having heard of privacy policies described in this way:

> Privacy policies are concise statements of what the operators of a Web site will do with information they collect from you, and how they promise to safeguard it.

Another 14% had heard of the term in this context, but had little or no idea what they are for or how they helped site visitors. 79% claimed to have at least some knowledge of Web site privacy policies.

Skepticism is widespread. Our respondents were overwhelmingly unimpressed with Web site privacy policies. Many used strong and colorful language to express their dissatisfaction *(e.g.,* "*CYA: cover your ass statements*", "*horse shit*", and "*as trustworthy as a politican's [sic] promises*").

In their open text responses, many of the respondents described one or more of the following weaknesses:

- privacy policies typically disclaim the sharing of information, rather than assuring its protection;
- the legal standing of privacy policies is not well known and is presumed to be very weak; and
- privacy policies are subject to change at any time, which is widely presumed to mean that site operators can, with impunity, ignore any promises they may have made to you simply by changing their policy.

A significant number of respondents appear to believe that the existence of a privacy policy automatically implies a promise of confidentiality when in fact they may disclaim it. One respondent described privacy policies this way:

> *"They are to protect any information you give to that particular site. You are protected from them giving out your personal information."*

This suggests that users may often be jumping to inappropriate conclusions when they see that a site has a privacy policy.

40% of respondents agreed that privacy policies help protect information after it has been collected, while 33% disagree. It is unclear how many of the respondents who agree also believe that the existence of a policy automatically implies a promise of confidentiality.

We trust you anyway. One result was quite surprising. We asked respondents to indicate their level of agreement with the following statements:

> If a Web site has a privacy policy, its operators have no choice but to respect it.
>
> A web site can violate its stated privacy policy, but most sites can be trusted to respect it.

Overall, only 9% agreed to some extent with the first statement. Two thirds (67%) disagreed. On the other hand, 44% agreed with the second statement, while only 18% disagreed. The pattern was quite evident when reading individual responses. One respondent after another would express deep skepticism with respect to nearly every question about privacy policies, but most appear to believe that sites in general can be trusted to be honorable. One respondent, who appeared to be knowledgeable in this domain, strongly disagreed with every single positive statement posed for any of the four technologies, with only one exception: he agreed that sites can be trusted to respect their policies. We could not find anything in the open text responses to shed light on this apparent contradiction.

Confidence. A large number of respondents, 73%, are not confident that they would know if a privacy policy was violated. Only 6% expressed confidence in this regard.

39% of respondents claim to be familiar with their browser's privacy policy features, but only 9% admitted to not knowing how to control them. 23% of respondents claim to have adjusted their browser's privacy settings to suit their personal preferences, and a further 7% have investigated the controls and are comfortable with the default settings. This means that nearly a third of our respondents are confident in their use of privacy preference features in their browser. We had expected this number to be substantially lower.

Awareness. 41% of respondents do not know if their browser has features relating to privacy policies, and a further 10% believe their browser has privacy features, but have never looked at them. Because these features are relatively new, having appeared only in the latest major versions of Internet Explorer, Netscape and Mozilla, we were not surprised to find that a majority of our respondents were unfamiliar with them.

Reliance on tools. 25% of our respondents reported that they would stop using some sites if the sites did not have privacy policies that were both understandable and acceptable. A further 6% would stop using some sites if the sites did not have privacy policies at all.

34% of our respondents reported that their decision to trust a site never depends on a stated privacy policy. Another 29% feel more comfortable with some sites because of the privacy policies they present, but would probably continue to use the site even if a written policy was not available.

5.2.4. Trust marks

We introduced trust marks in this way:

> Many Web pages display a *trust mark*. For example, you may have seen some of the trust marks below displayed on a Web page:

The images of the following five common trust marks were displayed below the introductory statement:

- TRUSTe Initiative Trust Seal
- BBB (Better Business Bureau) System Trust Seal
- VeriSign Secure Site Seal
- CPA WebTrust Electronic Commerce Seal

- ePublicEye Registered Safer Shopping Site Seal

Thirty-eight of our respondents (16%) reported never having heard of the term. Another 28% had heard of the term, but had little or no idea what they are for or how they help site visitors. 55% claimed to have at least some knowledge of Web site trust marks.

Some evidence that trust marks are trusted. 42% of respondents reported that they are more likely to trust a site that displays a trust mark, while only 19% said they were not more likely to trust it. Similarly, 49% of respondents reported that they are more likely to trust a site displaying a trust mark only if they recognize the trust mark program, while 12% reported that they are not more likely to trust it.

Only 32% of respondents reported that the trust they attribute to a site because of a trust mark is conditional upon its validation; 32% indicated that the trust they base on a trust mark is not conditional upon validation.

Other observations. Many respondents recognized that spoofing of a trust mark is a concern, even those who did not appear to have significant technical expertise. The following statement is typical of the way in which this was reported:

> *"Anyone can copy the graphic and put it on their site – it doesn't mean that the site is actually secure."*

One respondent appears to have confused the purpose of trust marks with the server authentication capabilities of HTTPS. He described trust marks as follows:

> *"third party companies which guarantee that the site i am communicating with is the actual site with whom communication is intended."*

Another's open text response was simply, "*provide a reliable source the [sic] the site's public key*".

One possible explanation for this confusion is that VeriSign's Secure Site Seal program intentionally couples its trust marks with the digital certificates it supplies to member sites. When one clicks on a VeriSign trust seal to validate it, the resulting information refers primarily to the authenticity of the site.

If this is indeed the real source of the confusion, then it may lead users to attribute server authentication properties to other trust mark programs that are not in fact connected with server authentication.

Several respondents believe that trust marks indicate that site security is managed by the trust mark company. This is a misconception, although it is not clear that it is a particularly dangerous one.

One respondent described trust marks in the following way:

> *"trustmarks are fancy buzz words used to placate the masses into making them seem trustworthy. Since their membership is pay only, the trust in them stops at the buck. About as trustworthy as CA's."*

This sentiment is reminiscent of Matt Blaze's remark that a commercial certificate authority will protect you from anyone whose money they refuse to take.

There is some evidence to suggest that people who know how to validate trust marks do not generally find the validation evidence compelling. Many of our respondents claim to be aware of the validation process, but are not confident that they could detect forgeries. We asked respondents to indicate their level of agreement with the following statement:

> I am confident that I would know if a trust mark displayed on a Web site was a forgery and not sanctioned by the trust mark program or company.

54% of respondents disagreed to some extent, while only 11% agreed.

Awareness of authenticity and validation. 12% of respondents reported being unaware that authenticity of trust marks is a concern (many of whom made this point explicitly in their open text responses). A further 23% were unaware that most trust marks can be validated by clicking on the graphic. 19% of respondents were aware that validation information for a trust mark can be viewed by clicking on the graphic, but had never done it. (It is unclear how many of these respondents had never had the opportunity.)

Confidence. As noted above, only 11% of respondents reported being confident that they could recognize a forged trust mark, while 54% reported a lack of confidence in this regard. Of the 68 respondents who reported being aware that validation information can be obtained by clicking on the graphic, 37 indicated that the validation process does not increase their trust. Only 31 of the 68 respondents are influenced by the validation.

Reliance on tools. Only 6% of our respondents reported that they would stop using some sites if the sites did not display trust marks. 64% either are never influenced by the presence of trust marks (35%), or feel more comfortable with some sites because they display trust marks, but would probably use the sites even if they did not (29%).

6. Summary and conclusions

We have described an online survey that explored typical Internet users' awareness and knowledge of specific technologies that relate to their security and privacy when using a Web browser to access the Internet. Over a four-month period, 237 individuals completed an online questionnaire. Respondents were predominately Canadian, with substantial numbers from the United Kingdom and the United States.

We had three Spanish and two Italian respondents (reported as country of residence). When reading through their responses to the opinion questions, all five stood out as following an unusual response pattern. Although the differences are not easy to characterize and we cannot draw specific conclusions from our data, it suggests that there may be cultural differences to be explored more deeply.

Because respondents were anonymous and self-selecting, the survey did not seek to precisely characterize the security awareness and knowledge of typical Internet users. Rather, it was used to identify potential misunderstandings and misconceptions, the most interesting of which are

summarized below. In many cases, it is clear that software designers could easily make different choices to avoid confusing or misleading some users. In others (e.g., skepticism about privacy policies), changes would be relatively difficult or expensive. In such cases, our observations may help identify areas where more precise quantification of the issues is needed.

6.1. Significant Findings

1. Users have tried to educate themselves regarding their security and privacy online, but with mixed results. It appears that many who have tried have had limited success because the subject is a difficult one in which technical subtleties are significant. This finding calls into question the assumption that a modest amount of education would be effective if only users were motivated to pursue it.
2. The term "secure Web site" is used in technical circles to refer to the use of SSL/TLS to secure the HTTP transport between a client and server. However, some users clearly interpret the term to mean that the site itself is secure in some assumed but unspecified way. Users who learn about the closed lock icon and other indicators of "secure sites" may therefore attribute security properties to the site itself whenever they see these cues in the browser. As a group, they tend to believe that the presumed security makes such sites more trustworthy for the purpose of conducting business.
3. Relatively few of our respondents agreed with the statement that secure sites provide assurance that the site with which they are communicating is authentic and not an impostor. It is unclear whether this indicates a lack of awareness of the server authentication component of the SSL/TLS protocols, or whether respondents disagreed for other reasons. Further investigation is necessary to resolve the matter.
4. Skepticism of privacy policies is widespread and our respondents expressed their views on this issue very strongly. Nevertheless, respondents generally seem prepared to trust that site operators will respect their stated policy even though they generally believe that the policies have no legal standing and can be changed at any time.
5. There is evidence of confusion between the roles of browser and Web server, especially with respect to the handling of cookies. In particular, many respondents confused cookie usage with browser-side caching and form-filling. We found evidence to suggest that the distinctions between the different types of information stored by browsers (cookies, bookmarks, cached pages and form data) are not clearly understood, and may lead to inappropriate conclusions about the impact of browser cookies.
6. Some respondents believed that trust marks provide some assurance of server authenticity. This confusion may arise from the tight coupling of trust marks with server certificates in VeriSign's Secure Site Seal program.

Acknowledgments

The authors would like to thank Michelle Anderson who helped build the online questionnaire web application during a co-op work term from January to April 2004.

References

[1] Branchaud, M. and Linn, J. Extended validation models in PKI: Alternatives and implications. In Sean Smith, editor, *Proceedings of the 1st Annual PKI Research Workshop*, pages 37-43. NIST, April 2002. Retrieved 23-Feb-2005 from http://www.cs.dartmouth.edu/~pki02/Branchaud/.

[2] Cheskin Research. Trust in the Wired Americas. July 2000. Retrieved 23-Feb-2005 from http://www.cheskin.com/p/ar.asp?mlid=7.

[3] Cranor, L.F., Arjula, M., and Guduru, P. Use of a P3P user agent by early adopters. In *Proceeding of the ACM workshop on Privacy in the Electronic Society*, pages 1-10. ACM Press, 2002.

[4] Dourish, P., Grinter, R.E., Dalal, B., Delgado de la Flor, J. and Joseph, M. Security Day-to-Day: User Strategies for Managing Security as an Everyday, Practical Problem. Technical Report UCI-ISR-03-5, Institute for Software Research, University of California, Irvine, June 2003.

[5] Ellison, C. and Schneier, B. Ten risks of PKI: What you're not being told about public key infrastructure. *Computer Security Journal*, 16(1):1-7, 2000.

[6] Fogg, B.J. Prominence-Interpretation Theory: Explaining How People Assess Credibility Online. In *Conference Extended Abstracts on Human Factors in Computer Systems*, pages 722-723, Fort Lauderdale, Florida, USA, April 5-10 2003. ACM Press.

[7] Friedman, B., Khan, P.H., Jr. and Howe, D.C. Trust online. *Communications of the ACM,* 43(12):34-40, 2000.

[8] Friedman, B., Nissenbaum, H., Hurley, D., Howe, D.C. and Felten, E. Users' Conceptions of Risks and Harms on the Web: A Comparative Study. In *Conference Extended Abstracts on Human Factors in Computer Systems,* pages 614-615, Minneapolis, Minnesota, USA, April 20-25, 2002. ACM Press.

[9] Friedman, B., Hurley, D., Howe, D.C., Felten, E. and Nissenbaum, H. Users' Conceptions of Web Security: A Comparative Study. In *Conference Extended Abstracts on Human Factors in Computer Systems,* pages 746-747, Minneapolis, Minnesota, USA, April 20-25, 2002. ACM Press.

[10] Garfinkel, S.L., Schiller, J.I., Nordlander, E., Margrave, D. and Miller, R.C. Views, Reactions and Impact of Digitally-Signed Mail in e-Commerce. To appear in *Proceedings of Financial Cryptography and Data Security,* 2005.

[11] Millett, L.I., Friedman, B. and Felten, E. Cookies and Web browser design: toward realizing informed consent online. In *Proceedings of the SIGCHI conference on Human factors in computing systems,* pages 46-52, 2001. ACM Press.

[12] Whitten, A. and Tygar, J.D. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *Proceedings of the Eighth USENIX Security Symposium (Security'99),* pages 169-183, 23-26 August 1999. USENIX Association.