Information Requested by Web Sites and Users' Comprehension of Privacy Policies

Robert W. Proctor
Purdue University; CERIAS
Dept. of Psychological Sciences
703 Third Street
West Lafayette, IN 47907
1-765-494-0784
proctor@psych.purdue.edu

M. Athar Ali
Purdue University
Dept. of Computer Technology
401 North Grant Street
West Lafayette, IN 47907
maali@purdue.edu

Kim-Phuong L. Vu
California State University, Long Beach
Department of Psychology
1250 Bellflower Blvd
Long Beach, CA 90840
1-562-985-5021
kvu8@csulb.edu

ABSTRACT
Two studies examined the information requested by different Web sites and users' comprehension of the sites' privacy policies. The first study analyzed Web sites from several categories with respect to the type and amount of personal information requested. It also examined the features included by the sites to assure users that the organization follows good privacy practices. Study 2 analyzed privacy policies of a subset of the Web sites identified for Study 1 in terms of readability and number of protection and vulnerability goals. We measured users' comprehension of four policies that were of equal readability but differed in the number of protection and vulnerability goals. Results showed that the Web sites varied in the amount of information collected and privacy practices. Also, users displayed poor comprehension of the Web privacy policies regardless of whether the policies were high or low in terms of protection and vulnerability goals.

1. INTRODUCTION
Issues relating to consumer privacy and the privacy policies of organizations are of central concern when interacting with the Web. Users are often asked to provide personal information such as social security numbers, credit card numbers, and medical history. The type of information elicited by a Web site may vary depending on the nature of the site. It seems appropriate for users to be asked to provide credit card numbers if they are purchasing items from an E-commerce site, but not if they are viewing airfares on a travel site. The willingness of users to provide certain pieces of personal and financial information depends on many factors, including preferences and prior experiences, as well as the purpose and design of the site requesting the information.

Ackerman, Cranor, and Reagle [1] conducted a survey regarding users' privacy concerns. They found several factors that were rated as highly important in determining whether users would provide requested information. These factors include:

a. whether their information will be shared with others,
b. whether the information will be used in an identifiable way,
c. the kind of information collected, and
d. the purpose for which the information is collected.

Between 30% and 50% of the respondents were concerned about whether the site posts a privacy policy, has a privacy seal, and discloses a date of retention policy. People indicated that they were more likely to provide personal information if they perceived it to be relevant to the Web site than if they did not.

A potential source of information for users is a Web site's privacy policy. Jensen and Potts [2] reported that 77% of Web sites post a privacy policy. However, a readability analysis showed that the most readable policies were at the 11.5 grade level, and the average readability was at 14 years of education. Although users stated that they were concerned about privacy policies, log files indicated that less than 1% actually viewed the privacy policy page. One reason for this is that considerable time and effort is required to locate, read, and analyze the policies [3].

2. STUDY 1: ANALYSIS OF WEB SITES
The goal of this study was to survey Web sites from several categories to determine the amount and type of information solicited from users when performing transactions with the sites. We also recorded whether each site posted privacy policies, included a privacy policy logo, and allowed users to select privacy preferences. We determined the percentage of sites within each category for each measure and examined the consistency of the recorded information within and between categories of Web sites.

2.1 Method
Privacy policies were selected from 42 Web sites and retrieved during the first week of February, 2005. They were stored for analysis to avoid modifications to the policies that could be made during the period of the study. The categories of sites were:

1. Financial institutions
2. Insurance companies
3. On-line pharmacies
4. Travel agencies and airlines
5. On-line game sites
6. Retail sites
7. Technology sites

We performed detailed analyses of the privacy- and security-related features of each site.

2.2 Results and Discussion
There was a difference in the amount of personal information requested by sites from the different categories, $F(6,35) = 15.47$, $p < .001$. Financial sites requested the most information from users, averaging a request for 30 pieces of personal information. The three most common pieces of information requested by the sites, regardless of category, were name (97%), postal address (88%),

and E-mail address (71%). Only a majority of pharmaceutical and retail sites asked for the user's credit card number. A user's home phone number was solicited not only by financial institutions, but by travel, retail, game, and technology sites as well.

Every site solicited users' personal information, but not all posted privacy policies. With the exception of on-line gaming, at least 80% of the Web sites in the other categories had links available to the sites' privacy policy information. Although a large amount of private information is requested by the Web sites examined, considerable differences exist within and between categories with respect to the type and amount of information a user is asked to provide. Given the competitive nature of the Web and that users have the option of leaving a site without completing a transaction, an organization should promote user trust by ensuring that their site does not ask for unnecessary information.

3. STUDY 2: POLICY COMPREHENSION EXPERIMENT

This study investigated users' comprehension of privacy policies and perceptions of the security provided by them. Policies from 25 sites in the pharmacy, retail, banks, and insurance categories from Study 1 were selected for content analysis. Analyses were performed to determine the number of protection and vulnerability goals [4], as well as each policy's readability.

3.1 Methods

20 college students, who were at the 13-year reading level at which the average policy is written, participated. All were experienced with Web-based transactions. Four privacy policies that were similar in reading level were evaluated, with two being low in protection and vulnerability goals and two being high. The names of the companies were removed to eliminate the possibility that previous experience with any of the organizations could influence users' performance and preference judgments.

Participants read each of the four privacy policies and answered questions relating to the policies. They also reported their preferences concerning specific aspects of each of the policies. Each subject viewed all four policies, with the order in which the policies were viewed counterbalanced across subjects.

3.2 Results and Discussion

Time to read each policy was proportional to the word count (r = 0.63), indicating that subjects were in fact reading the policies. Overall, users showed poor comprehension of the privacy policies (see Table 1). There was a main effect of policy, $F(3,57) = 5.56$, $p = .002$, with performance being poorest for policy D, which was low on both protection and vulnerability goals.

The two policies that were low on protection and vulnerability goals (polices A and D) were judged by the majority of participants as not providing measures beyond those required by law. Policies B and C, which had a higher number of goals, were judged to be better than policies A and D at assuring information protection. However, when asked if these policies provided extra measures than required by law, 40% and 35% of users said "no" for the respective policies. These percentages, though lower than those for policies A and D, are still substantial.

Web site type	Identifier	% correct	Web site category
Short length (510 words) & low goals	A	58.7	Bank
Long length (1628 words) & high goals	B	58.9	Bank
Long length (2100 words) & high goals	C	47.5	Retail
Medium length (1306 words) & low goals	D	36.3	Retail

Table 1: Policy Comprehension Test Results

4. SUMMARY AND CONCLUSIONS

Study 1 showed that different types and amounts of information were requested by Web sites within and between categories. Generally, sites requested a lot of personal information, but not all sites followed best practices of posting privacy policies and privacy logos. Study 2 showed that people have poor comprehension of Web privacy policies and are hesitant to provide information to sites with short policies. This hesitancy could be due at least in part to the smaller number of protection goals that are contained in short policies than in longer ones.

The present studies show that implementation of privacy policies in the current state is not perfect. Best practices, such as including privacy logos and having highly readable privacy policies that can assure people of privacy protection, are often not followed. Given that people tend to be more assured by longer privacy policies than by shorter ones, they are more likely to give their information to sites that specify many protection goals compared to those that do not. The privacy policies that have high protection and vulnerability goals are perceived by users as providing better security measures but to be more repetitive than those that have low protection and vulnerability goals. Future studies should examine how to organize and present privacy information in a manner that is easier for users to comprehend.

5. ACKNOWLEDGMENTS

This research was supported by NSF ITR Cybertrust grant #0430274.

6. REFERENCES

[1] Ackerman, M. S., Cranor, L. F., & Reagle, J. (1999). Privacy in E-commerce: Examining user scenarios and privacy preferences. *E-commerce* 99 (pp. 1-8). New York: ACM.

[2] Jensen, C., & Potts, J. (2004). Privacy policies as decision-making tools: An Evaluation of online privacy notices. *CHI 2004, 6,* 471-478.

[3] Byers, S., Cranor, L. F., & Kormann, D. (2003). Automated analysis of P3P-enabbled Web sites. *ICES* 2003 (pp. 326-338). Pittsburgh, PA.

[4] Antón, A.I., Earp, J.B., He, Q., Stufflebeam, W., Bolchini, D., & Jensen, C. (2004). The lack of clarity in financial privacy policies and the need for standardization. *IEEE Security and Privacy*, *2(2)*, Mar-Apr, 36-45.