# Citizen's Attitudes about Privacy While Accessing Government and Private Websites: Results of an Online Study

Salvatore J. Stolfo[1], Eric Johnson[2], Tomislav Pavlicic[3], Stephen Jan[1]

sal@cs.columbia.edu, ejj3@columbia.edu, tomislav@paradox.psych.columbia.edu, sj178@columbia.edu

Columbia University

Digital Government Research Center[4]

New York, NY 10027

**Abstract**

This paper reports the results of an investigation on citizens' attitudes and concerns regarding privacy and security on the Web, in general, and on the government websites they may visit, in particular. We examine to what extent those concerns can be alleviated by using a Secure Private Portal that protects citizen's personally identifying information when accessing government websites. The research project had two main goals: (a) to develop a comprehensive psychological instrument to assess citizens' attitudes and concerns regarding privacy and security on the Web; (b) to test the impact a Secure Private Portal may have on those concerns and on the way citizens use Government Websites. In order to accomplish these goals researchers from Columbia Business School and from Columbia departments of Computer Science and Psychology, developed and ran a web-based survey. Participants were recruited using online advertising through Google.com and provided their responses on the web. Early analyses of the results indicate a very high level of citizens' concerns regarding privacy and security of their personal data. Some of the concerns can appropriately be addressed only by fundamental policy changes. Furthermore, the results suggest that citizens perceive those sites which use secure portals as much safer and are more likely to visit them again. The results may indicate a new strategy for the presentation and design of government websites.

## 1 Introduction

Government agencies are challenged with a major social and political issue, the erosion of ***citizen privacy*** due to the use of electronic communication and data services, especially the broad-based use of the Internet. Many agencies exist to provide publicly available sources of data about the citizenry for study, analysis, and ultimately to develop policy and legislation of benefit to the nation and the well being of the citizenry. However, these public sources of subject data are now the target of sophisticated and broadly available technologies that may easily reveal the true Personal Identifiable Information (PII) of the citizens that are the subject data. Public data enriched with other commercially available sources of data may serve to reveal many private aspects of any citizen's private life[5]. The fedstats.gov website provides direct access to many relevant agencies and sources of data that may be linked together to assist in the process of disclosing PII.

---

[1] Department of Computer Science, Fu Foundation School of Engineering and Applied Science.
[2] Department of Marketing, Columbia School of Business.
[3] Department of Psychology, Columbia Graduate School of Arts and Sciences.
[4] Supported by an NSF Digital Government program SGER grant, number 0140304-0027078000 – 12/31/2002.
[5] See http://www.gao.gov/special.pubs/pubshort.htm.

There is another significant problem faced by nearly all government agencies and which is not fully appreciated by those agencies and the public, namely, the inadvertent disclosure of, and needless gathering of PII of citizens who access government data and websites on the internet. Use of the Internet to acquire data from the citizenry more efficiently and at lower cost is a natural approach being considered by many agencies. However, such an approach creates deep concerns about the authenticity of the respondents who may answer online, and the opportunity this may create for fraud and for the invasion of citizen privacy by the inadvertent leakage of PII (private information may be revealed to unauthorized third parties).

Users and ordinary citizens who access government websites are subject to other privacy risks, which may put government agencies in a mode of operation that is not compliant with executive orders and OMG guidelines[6]. Users often search publicly available data sets for information that is of personal interest to them. Even if government websites do not track their users, unrelated third parties can capture the activity at the government website, including click-stream data and submitted queries and results and infer information about the users as well as the data sets they access. The social and political issues of gathering PII are obviously very important when considering the practices of government agencies. Although the Internet represents an extremely efficient distribution channel (between government to citizen), it's technical implementation also makes it an extremely efficient surveillance system. The response by many public advocacy groups, and other organizations, to this realization has prompted intense scrutiny of government web site practices.

The problem, however, is quite a bit more complex. Several well-publicized cases of espionage and hacker penetration of government computer network systems have created deep concerns about the security practices of government agencies. Directives have been issued by the executive branch of government requiring agencies to improve their security practices by operating sophisticated security technologies to protect the critical government network infrastructure. One of the component security technologies that are now being fielded is Intrusion detection systems[7]. These software systems continuously observe network traffic in order to detect known misuses or attempts to penetrate computer servers by hackers with malicious intent. Some particular classes of network attack (Distributed Denial of Service, for example) can be detected by observing the frequency of service requests from particular IP addresses. Hence, the security systems protecting government websites are in fact gathering IP addresses of citizens who visit those websites for security reasons. In fact, IP addresses reveal PII more directly than cookies.

The issue of balancing security of the government infrastructure with the security of the citizenry who use that infrastructure is subtle and has not been fully and openly discussed and studied. How might a government website protect itself from malicious hackers, while concurrently protecting the privacy of citizens who visit those websites?

We believe technical solutions to ensure the privacy of citizens when visiting websites may exist. The essence of the proposed solution is to provide privacy to users via ***Secure Private Portals*** to government websites. These portals serve the role of securing user's identities by generating on demand proxy identities on behalf of users that are presented to the government website. Furthermore, the citizen-to-government communication interaction can be routinely secured using widely deployed standard encryption technology. Concurrently, these private portals may also act as a security sentry to government web sites adding an additional layer of protection of the government's technical infrastructure.

---

[6] See www.plainlanguage.gov/example/other/privacy%20policy.htm.

[7] See, for example, www.cs.columbia.edu/ids/library/index.html.

The essence of the technical solution involves two key elements, a "trusted third party" that serves as custodian of (online) PII and a proxy server system that creates **one-time use proxy identities** on demand (with possibly a map between proxy identity and real PII; the map is held by the trusted third party). Proxy servers for anonymous web browsing have been available on the Internet for several years. For example, www.anonymizer.com is perhaps one of the first such anonymizing proxies broadly available that shield client IP addresses. The client IP address (the real PII) is proxied with the anonymizer.com IP address (the proxy PII). However, when visiting a website via the anonymizing proxy, users who provide their real email address to that website have gained nothing; their true PII is revealed. Hence, total privacy cannot be achieved with only limited shielding of partial identity information. Rather, an end-to-end proxy shield is required that proxies all PII, including email addresses, and the basics of offline PII, name, address and phone number.

A secure private portal was implemented for this study to expose users to different browsing experiences, and to see how users would react to them. The portal site has two components. The first part of the website serves as a mechanism to point users to government websites (a very large list of sites was composed after an exhaustive search). This component of the site has two modes of operation - secure and insecure. Under the insecure mode, the portal allows the user to browse the website without any interference from the portal. Under the secure mode the portal serves as a secure shield that anonymizes any transaction a user makes through the portal. A secondary service provided by the portal is a service that generates proxy email addresses. A user may enter their real email address and the proxy email service will generate a temporary proxy email address that forwards mail to the real email address. This enables users to keep their email address private by filling forms with the proxy email address when requesting information at a website. The portal is implemented using a combination of JAVA, PHP, and HTML. The other major component implemented for this study was the survey instrument that responders experienced after they completed their browsing[8].

In summary, the likelihood of revealing PII of users who may access government websites is now very high. This access risk must be assessed in order to comply with the executive orders and OMB guidelines that prohibit government websites from tracking users, and needlessly gathering the online PII of citizens who access these public sources of information and services. Indeed, at present there has not been a formal study of citizen's perceptions and their preferences about their privacy and security when visiting government websites. It is very important that the fundamental relationship of trust between government and citizen when interacting on the web be better understood. This paper is a first step towards this understanding. To explore this area more deeply, a collaboration was formed between computer scientists, political and social scientists working with government agencies and a panel of anonymous citizens who were the subject of a formal research study to assess their attitudes. We detail the particular method used to implement a survey instrument, and the anonymous panel of citizen responders to that survey. The results of a portion of this survey and a critical finding related to secure private portals are detailed in the next sections[9].

## 2 Methods

### 2.1 Participants/Respondents

The respondents to the survey were one-hundred-and-ninety-five U.S. residents age 18 and older who responded to the advertisement posted on Google.com and who completed the survey. Respondents were paid for their participation.

[8] The implementation details will be fully described in a subsequent report.
[9] A future report will be more expansive about all of our findings than space allows here.

### 2.2 Materials

Materials consisted of an online questionnaire constructed to measure privacy attitudes. Most of the items on the questionnaire were adapted from Bellman at al (2001) and Smith, Burke and Millberg (1996). Several items were adapted from Business Week / Harris Poll (March 20, 2000) survey and from Pew Internet and American Life Project (February 2001) as well as from GVU Georgia Tech surveys. The questionnaire was specifically concerned with participants' attitudes towards the amount and accuracy of collected information as well as the security of that information from access and disclosure risks. In addition the questionnaire contained a set of questions intended to assess the respondent's knowledge of privacy technology, and a set of demographic questions. Finally, some questions required subjects to assess the security offered by the Government websites they visited as a part of the survey.

### 2.3 Procedure

The study was advertised on the Google.com. Whenever a Google.com user performed a search which contained "gov" as a part of the search string the following advertisement appeared:

Visiting Gov. Sites?

Worried about your privacy?

Answer our survey and win $10!

People who responded to the ad were randomly assigned either to the experimental or the control condition. Participants in either group were offered an extensive list of federal, state and local government websites and asked to visit as many of them as they wanted. However, participants in the experimental group were also informed that their identity and personal identifiable information would be shielded by a Secure Private Portal. In addition, the instructions for the experimental group contained, in lay terms, a brief explanation of the features of the secure portal. Once participants finished browsing government websites they were asked to answer the questionnaire. To counteract possible effects of the presentation order two forms (A and B) of the questionnaire were created by reversing the order of potentially sensitizing items. One half of the participants in each experimental group received Form A and the other half Form B. Subsequent analyses did not reveal any significant differences between the responses to each of the two forms and the results were pooled for all the analyses reported in this paper.

## 3 Analysis and Results of the Survey

Our sample, while not a result of probability sampling of all users of government Internet sites, was similar in many ways to what we know about the U.S. Internet population as a whole. They reported being 49% female and 47% male, 80% Caucasian, 6% African-American and 52% were married. They came from 40 different states. We also gathered information about why they were visiting a government site: Twenty-seven percent were looking for a form or application, and 66% were looking for contact or location information.

We examine three questions in turn: (1) how concerned and knowledgeable are these respondents about security and privacy in general, (2) how does this concern and knowledge vary across government and commercial sites, and (3) what is the influence of a privacy portal on these concerns and knowledge?

As in past research, there is significant concern about privacy. For example 66% of our respondents agree or strongly agree with the statement that they are "usually bothered when a web site asks me for personal information" and 80% either agree or strongly agree that they "think twice before providing personal information." As in our past research, a strong majority of (70%) respondents strongly agree with the statement that a web site should ask for explicit permission before sharing the information provided. This

similarity to past research (Bellman et al., 2001) suggests that neither time or sample differences have diminished the reported concern about privacy.

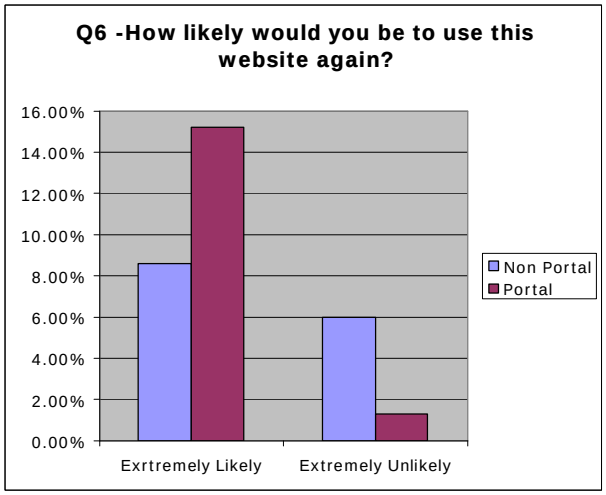
We examined the reported concern of our respondents about security of communication for both government and commercial web sites. In both cases, there was significant concern. Sixty-six percent of respondents were very or extremely concerned about third party monitoring of government sites with the corresponding percentage for commercial sites being 68%. Interestingly, there was less concern with the security of transactions, and it seemed greater for commercial sites (57%) than for government sites (21%). This might reflect the current smaller frequency of financial transactions with government sites.

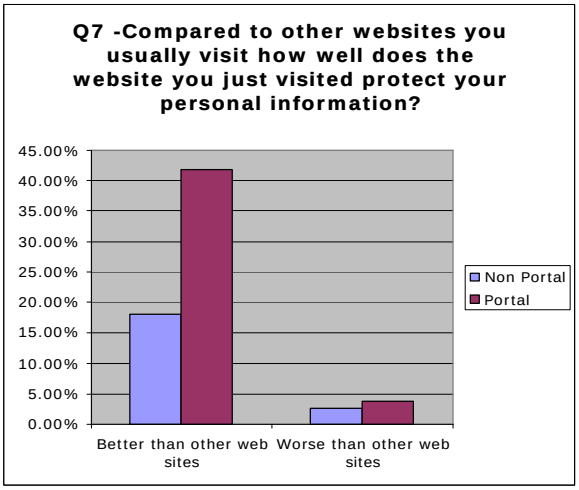
Finally we turn to the evaluation of the portal. Remember that respondents saw these questions after they had conducted their business with the government site, and that roughly half of them, randomly assigned had done so through the portal. For purposes of this report, we concentrate on two questions, both referring to the visit that they had just made. (Plots of the answers to two of the survey questions pertaining to the portal concept appear below.)

First, we asked respondents how this site protected their personal information compared to other websites. Here, the portal seemed effective: Forty-two percent of the portal users answered slightly or much better than other websites, compared to 18% of the non-portal users. We must note, however, that many respondents 41% of portal users, and 51% of non-portal users were unsure, and answered that they did not know. While the portal apparently increased feelings of security and reduced uncertainty, there seems to be room for improvement in educating users about the portal. Most importantly, the portal seemed to increase the likelihood that they would return: While 38% of the non-portal users said it was very or extremely likely that they would return to this site, 47% of portal uses said it was very or extremely likely that they would return. To test the statistical significance of all portal vs. non-portal comparisons, we conducted a chi-square test on the frequencies of responses. For both the protection and return likelihood measures, the results are at least marginally significant ($p < .001$ and .075, respectively).

## 4. Concluding Remarks

Thousands of anonymous web surfers visiting government websites were exposed to our research project, with about 200 individuals participating by completing and submitting a survey instrument. Early analyses of the results indicate a very high level of citizens' concerns regarding privacy and security of their personal data. Some of the concerns can appropriately be addressed only by fundamental policy changes. Furthermore, the results suggest that citizens perceive those sites which use secure portals as much safer and are more likely to visit them again. The results may indicate a new strategy for the presentation and design of government websites. The results reported here have focused primarily on a finding related to the concept of a secure private portal. A more extensive report exploring a number of other findings of the survey is in preparation and will be published at a later time.




## 5 Bibliography and Online References


1. OMB Guidelines, see www.plainlanguage.gov/example/other/privacy%20policy.htm.
2. Pew Internet and American Life report, see see http://www.newsbytes.com/news/01/166196.html.
3. "Record Linkage and Privacy", a report from the US General Accounting Office, GAO-01-126SP available at http://www.gao.gov/special.pubs/pubshort.htm.
4. IDS publications, see http://www.cs.columbia.edu/ids/library/index.html
5. Bellman, S., Lohse, G. and Johnson, E., Predictors of Online Buying, **Communications of the ACM.** Available at http://cebiz.org/Papers/wvtm1.pdf
6. Johnson, E. J., Bellman, S, and Lohse, G. Defaults, Framing and Privacy: Why Opting in Does Not Equal Opting Out. In press, **Marketing Letters.** Available at http://cebiz.org/Papers/optinoptout%20submit.PDF.
7. Bellman, S. Johnson, E. and Lohse, G., To Opt-In or Opt-Out: It Depends on the Question. **Communications of the ACM,** available at http://cebiz.org/Papers/Published/Opt-In_Opt-Out CACM.PDF.
8. Bellman, S. Johnson, E., Kobrin, S. & Lohse, G. Cultural and Regional Influences on Concerns about Internet Privacy and Security, available at http://cebiz.org/My%20PDF%20Papers/global%20concerns%20about%20privacy.pdf
9. Smith, J., Milberg, S., & Burke, S. Information Privacy: Measuring Individual's Concerns About Organizational Practices, MIS Quarterly, 1996, 167-106.
10. Lohse, G., Bellman, S, and Johnson, E., Consumer Buying Behavior on the Internet: Findings from Panel Data. **Journal of Interactive Marketing**, 2000, Available at: http://cebiz.org/Papers/99wvtm2.pdf.