An Investigation into User Perceptions of Privacy and Trust and their Real-World Practices

R.Stockton and S.Cunningham

Creative and Applied Research for the Digital Society (CARDS), Glyndŵr University, North Wales, UK
e-mail: {r.stockton|s.cunningham}@glyndwr.ac.uk

Abstract

Internet security is a well-researched and understood topic. However, the concept of online privacy is a less well-defined issue, particularly with the advent of mass email and social networking interactions. This paper presents results from a user survey to determine perceptions of privacy and trust in online environments. Results suggest that users have often-contradictory views of their privacy in the online world, but that the content of personal messages are perceived as the most sensitive data exchanged online. There is recognition of a number of security mechanisms available in the web domain. We go on to investigate personal messaging security in email. This is done by an investigation into the adoption of Transport Layer Security (TLS) encryption facilities in Internet mail servers, split into a pilot study and larger-scale piece of work. Results show that the majority of email servers sampled utilise TLS and that this majority is only somewhat more than 60% of all servers tested.

Keywords

Privacy; trust, security, encryption.

1. Introduction

This paper provides an initial investigation into user perceptions of privacy and trust in the online world and undertakes an initial examination of technologies that may be deployed to secure the information deemed most at-risk by user groups. In doing so, we seek to map out the current landscape of Internet privacy and to demonstrate that online privacy is an area in its infancy and one that needs significant further research to determine models of best practice for policymakers and security technologists.

2. Background

2.1. Online Privacy

Kemp and Moore (2007) discuss privacy as being a difficult to define, that it has quite clearly been eroded in the USA with legislation such as the PATRIOT act since the terrorist attacks of September 2011, and a similar situation is in place in the UK. The work represents Solove's (2002) six conceptions of privacy: 1) the right to be let alone; 2) limited access to the self; 3) secrecy; 4) control of personal information; 5) personhood; 6) intimacy; and adds a seventh term: 7) privacy as a cluster concept. These definitions have their problems, but are all clearly issues in the online society.

Warren (2002) analyses the legal context, provision of the UK Data Protection Act of 1998 and provides an evaluation of the approaches that have been taken by organizations to action the DPA requirements. The body of the work in the paper looks at the approaches taken by organizations to implement the DPA. The researcher uses three methods to examine the organizations. These are: a questionnaire survey of 14 organizations; 30 interviews of experts from across government and public sector organizations; and case studies compiled from in-depth interviews with employees from organizations including health, police and education. Warren finds that the scope of the right to privacy under this new legislation remains untested and contested by employer bodies. The paper identifies a variation in investment in compliance with the DPA from £75 to £10,000 per year and that awareness raising and training throughout an organization is important to comply with this legislation.

Yee (2006) states that privacy protection approaches are in early stages of research, treating the problem as an access protection one through technology and rights management. Yee argues it is important to measure the trust that the service user can have in the web service, as without trust, the service will not be used. Yee references the Goldberg *at al.* (1997) definition of privacy: "*privacy refers to the ability of individuals to control the collection, retention, and distribution of information about themselves*." Yee takes this statement and outlines the following definitions: 1) privacy refers to the ability of individuals to control the collection, use, retention, and distribution of information about themselves; 2) a service's protection of user privacy refers to the service's use of provisions to give the user control over the service's collection, retention, and distribution of information about the user; and 3) a measure of a service's protection of user privacy is a numerical value that indicates the degree of the user's control (or some aspect of that control) over the service's collection, retention, and distribution of information about the user.

Yee states there must be an understanding between the service provider and the service user as defined by a number of rules. Without consequences by a legal or governance framework then the trust by the users towards the service provider will not be a strong. Yee outlines some basic provisions that need to be in place and measured to control the trust and privacy of the service user's information and states that the list is not exhaustive:

- Use of a privacy policy that automatically ensures that the user's privacy policy is not violated;

- Use of a cryptographically secure log (this log can be later inspected to check for policy violations) to record each provider action involving the user's private data;

- Use of employee background checks when they are hired to try to exclude dishonest people from the provider's organization;

- Use of reputation mechanisms to record and indicate the past performance of the provider in terms of integrity (e.g. Better Business Bureau);

- Use of seals of approval that attest to the fact that the provider has undergone and passed rigorous inspections of its business processes.

Yee concludes with two outcomes, relevant to the research undertaken here. The measures serve at least two important functions: 1) they help the consumer to choose services that are more effective at protecting privacy, and 2) they let web service developers or managers know if more countermeasures are needed to achieve a predefined level of privacy protection effectiveness.

Kosa (2010) explores the current state of thinking in privacy related to computer systems. Work on formalizing trust in computer systems dates back to the mid-nineties, in contrast, however, work on formalizing privacy in computer systems is in its infancy. Kosa's work notes that privacy, unlike trust, is legislated [in Canada]. However, in regard of trust, the following statement is made:

"*Trust, on the other hand, is the Wild West; almost anything goes. Early attempts at formalizing privacy in computer systems have been largely restricted to P3P initiatives and other policy developments.*"

Kosa explains that privacy is historically treated as an emotional and ethical concept. Whereas trust is more about the acceptance of risk to support action. In the realm of computer systems computational trust tends to be about transactional reliability and authenticity. For the concepts of trust and privacy to be enabled at the technical level for computer systems to operate with, both need to be formalized.

2.2. Secure Messaging

Farrell (2009) discusses why users are not routinely encrypting their e-mail. It outlines that e-mail sent between a Message User Agent (MUA) and Message Transfer Agent (MTA) is normally unencrypted. Encryption systems for e-mail include: OpenPGP which can be installed as a plug in to most e-mail clients; and Secure Multipurpose Internet Mail Extensions (S/MIME), which is built into most e-mail clients. These protocols have been around for a long time. S/MIME has been implemented into MS Outlook since 2000 and Open PGP has been in existence since 1997. These encryption products can provide two services: origin authenticity and encryption. The paper outlines that to enable the use of any of these systems the user has to first arrange to exchange encryption keys between parties to allow the system to operate. This hurdle for the end user is the problem as they either don't know how to do this, don't know the option exists, or finds the steps too hard. Both systems also insist that users include proof of who you are to enable the origin authenticity element of the process.

Robison (2012) outlines a secure overlay feature implemented in a Javascript Bookmarklet. The overlay is used to provide an encrypted chat session in Facebook using Facebook chat as the communications channel and the Bookmarklet acting as the input and output window. All communications are encrypted and not viewable by Facebook. The paper explains the usability of the system and how it addresses the key problem with security in ease of use. A trail of the software showed that over 50% of people were able to use the system with relative ease. The paper describes a

survey conducted with a small sample size of 65 users that finds that people are unaware of any privacy issues associated with using online chat but assume that the phone and e-mail are more secure. This is a good clear piece of work showing how an overlay system can be used to hide communications across an untrusted system such as Facebook in this example. More work would be valuable in this area to include a generic system that could overlay many different web applications to secure the messages that pass through them.

Even though privacy is becoming a key requirement, frameworks that consider privacy in a comprehensive way are still missing. Most of the work in the literature focuses on a few aspects of privacy only. Moreover, much of the research has been devoted to anonymity metrics for privacy-preserving micro data releasing.

Colombo (2012) presents a model-based framework MAPaS. The MAPaS model is intended for developers of information systems to be able to create a privacy and access relationship model for the system under development. This model describes the information types in a schema with roles, skills and requirements of users of the system under development. Building access and information relationship rules in the system the developer is able to analyze and test the design for privacy weaknesses. Asking a number of questions of the model, it is possible to examine if the information and role allocation and access design is robust and meets the developers' intentions. An example given in the paper is attempting to allocate an access role to an inappropriate member of staff. If the privacy model description of the information system is correctly configured then the inappropriate staff member should be denied the inappropriate access role. If this is not the case the system logic can be amended. Using such as system over much iteration can assist a system developer to understand how the privacy logic of a system will respond before the expensive phase of actual systems development and testing.

3. Investigating User Perceptions of Privacy & Protection

3.1. Methodology

An online survey was devised to investigate the perceptions that users have of trust, security and privacy in their online activities. The survey was sent to all staff in a University as well as several JISC mailing lists including UCISA (Universities and Colleges Information Systems Association) and HEWIT (Higher Education Wales Information Technology). As such, it is acknowledged that the majority of respondents are people employed in the UK higher education sector. A total of 218 responses were received. A cross section of ages was represented in the survey with results mainly coming from the 35-50 age group category. When working with Chi-Square statistics, in our analysis, for less than 2 degrees of freedom, we apply Yates' correction.

3.2. Survey Results

Initial questions focused around the typical application uses of users surveyed. E-mail was the most used Internet application, followed closely by online web searching. Online shopping and banking were used extensively. There was a spread

of other applications being used by significant percentages of respondents. To determine the area where users felt their privacy is most vulnerable, they were each asked to select the three areas that concerned them most, from a pre-defined list of options. A summary of the responses is shown in Figure .

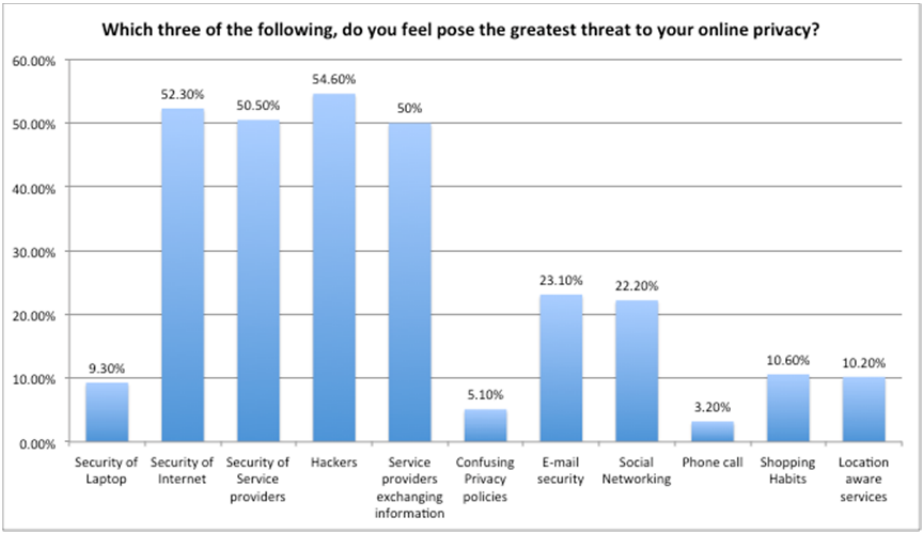

Figure 1: Threats to Online Privacy (*N*=216)

Hackers were the main concern at 54.6% followed closely by Security of Internet (52.3%) and Security of Service Providers which you send information to (50.5%). Curiously e-mail security (23.1%) and social networking (22.2%) were lower down the concerns. Shopping Habits (10.6%), Location aware services (10.2%) Laptop security (9.3%) and Phone calls (3.2%) had low concerns. Respondants seem to put greater importance on security and malicious threats from third parties than on the services providers not complying with privacy policies or having suitable or relevant privacy policies in place.

The UK Data Protection Act (1998) makes provision for the way in which organisations are entitled to store and use individuals' information. Part of the Act entitles individuals with access to this information, partly to ensure that it remains current and correct. However, we hypothesise that correctness of information does not equate to a sense of privacy or trust, since it remains in the possession of third party, whose uses of the information may not be as transparent to the end user. As such, we posed the question: "*As long as the information stored about you is correct then you have nothing to be concerned about". Do you agree with this statement?*". The results were that 86.2% responded negatively, indicating that the significant majority disagree, $\chi^2_{Yates}(1, N=218)=113.07$ $p<0.01$. It is clear that there are concerns on storage of information, which would indicate a lack of trust between users and data controllers.

In terms of users being aware of the privacy polices and terms of use, we asked participants "*Have you read the privacy policies provided by web sites you use and*

the software you install?". 41.3% responded Yes and 57.8% responding No (N=218) it is clear that a larger number than expected do read the privacy statements provided by the sites which shows they are interested in privacy online. This is not to say that they were satisfied with what they found in those statements. As a follow-up to this question, we asked the participants "*Have you refrained from using a service due to a requirement to provide them with what you consider to be personal or private information? Perhaps during a sign up stage or during a purchase.*" The results here were extremely significant, with 92.2% of respondents replying that they had refrained from using a service due to a requirement to provide what they considered to be personal or private information, $\chi^2_{Yates}(1, N=218)=153.62$ $p<0.01$.

To follow-on, it was decided to determine how frequently users divulge what they consider to be private information, in return for receiving a product of service, which they might otherwise prefer to not to share. To this end, a specific question employed was "*Have you revealed information about yourself to a service provider which you would have preferred not to, but did so to access the service?*". The breakdown of responses is shown in Figure 2. The results do not indicate specifically whether a majority of users do, or do not, reveal private information in return for a service, however, the low percentage of users reporting in the Don't Know category, indicates that the majority of users do manage to make an informed decision about the private information they may or may not reveal, $\chi^2(2, N=218)=69.61$ $p<0.01$.

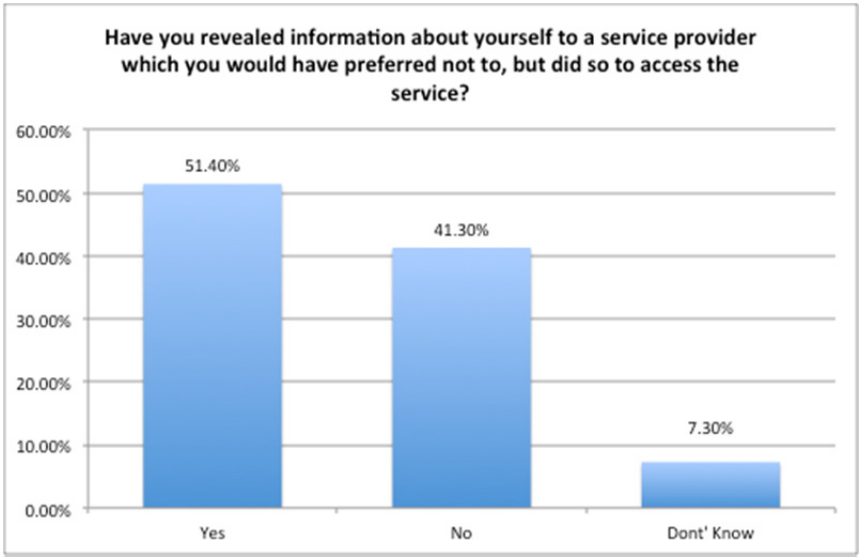

Figure 2: Revealing Private Information to Receive a Service (N=218)

To obtain an insight into technical features that users perceive as providing them with security features, participants were asked to select a technical security feature that would provide them with the most confidence in using an online service. The results can be seen in Figure 3, indicating a significant confidence amongst participants in in the use of Secure Socket Layer (SSL) χ^2 (3, N=218)=269.93 $p<0.01$. Results suggest that users value the recognized presence of SSL services

almost universally across online services, to provide a measure of confidence in their transactions, even more so that the company brand or other descriptive measures.

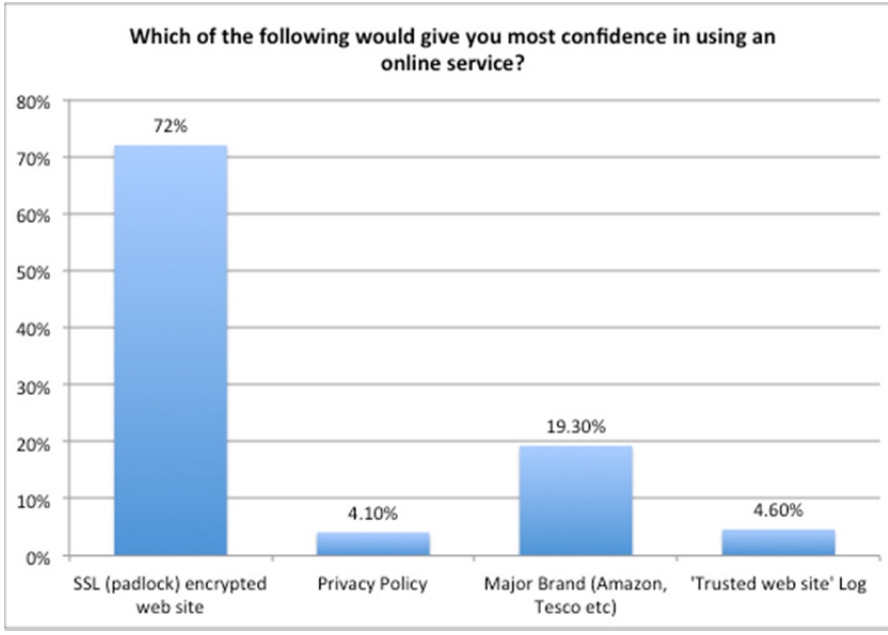

Figure 3: Technologies that give Confidence in Services (*N*=218)

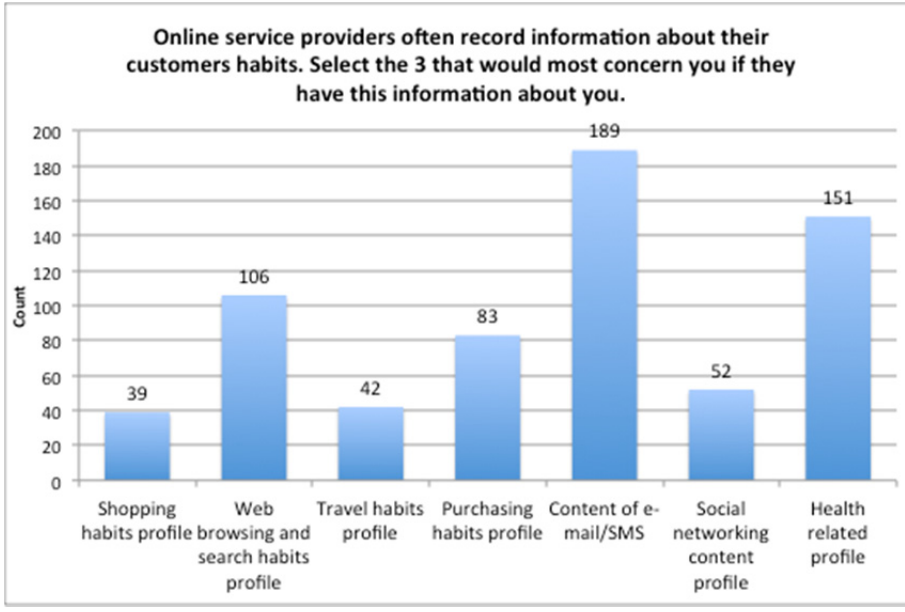

Figure 4: Most Sensitive Online Information (*N*=218)

Finally, participants were surveyed to determine the type of information that they perceived to be the most sensitive and wished for third parties to not have access to. The results of this investigation are shown in Figure 4, participants were asked to select the three most sensitive types of information.

Perhaps unsurprisingly, the contents of personal communications, email and SMS messaging, are in a significant majority, χ^2 (6, N=662)=211.79 $p<0.01$. More interestingly, is that this information appears to be slightly more sensitive to users than their medical or health data. The more open and generic information that might be disclosed in web browsing and online searching is considered more critical than more focused online purposes, such as shopping, travel and social networking. This suggests that users may have more of an awareness that their data is tracked in these more structured and enclosed interactions, but less so in general purpose Internet use. It is also possible that users regard data already in these systems to be 'available' as they have provided it knowingly and with a degree of informed consent.

4. Technical Measures for Privacy

Using Transport Layer Security (TLS) removes the need for the end user to configure encryption and simplifies adding this layer of security. The mail system administrator must configure and enable the TLS encryption at the mail server level. This system also relies on the receiving server being correctly configured to receive the encrypted TLS communications. When correctly configured at both ends of the communication TLS provides encrypted transport between the e-mail server across the Internet to the point of the receiving e-mail server. This prevents eavesdropping at any point in between including on a local LAN or the Internet. This system does not protect the end user from interception within the mail system be a mail administrator. To this end, we seek to answer the following question: *how widespread is TLS encryption implemented and is it correctly configured?*

4.1. Pilot Study: Methodology

Item	Description
MX count	The number of mail exchange servers that were advertised in the DNS record for the e-mail address domain
Connected	That the e-mail server could be contacted during the test
TLS Available	Test that TLS is advertised as available on the e-mail server
Cert OK	Check that the digital certificate is valid, had not expired, not been revoked and is designated for this domain
Negotiated TLS	Test if the TLS negotiation completes successfully

Table 1: TLS Analysis Points

As pilot exercise a sample that used 1736 MX points was conducted to investigate this question. Using a number of e-mail address harvesting tools to collect e-mail addresses from web pages and a number of consenting individuals' mailboxes a TLS test engine was used to connect to the mail server for each of the e-mail addresses and conduct the TLS protocol setup procedure. The test engine recorded which stages of the TLS protocol where initiated correctly for e-mail mail server and if the

server mandated TLS and required a trusted digital certificate or would allow for an untrusted digital certificate to be used. The original set of e-mail addresses contained many duplicates of domain names and these were removed so only one e-mail address for each domain name was tested. The testing collected data, described in Table 1, related to each email address

4.2. Pilot Study: Results

It was found that just over half of the domains tested (54%) had mail servers with TLS enabled. This does not mean, however, that half of the mail communicated between these servers will be encrypted using TLS as both end points are required to have TLS for the protocol to take effect. Assuming that there was an even distribution of communication between domains, this would show less than a quarter of the traffic between these servers would be using TLS.

The average number of MX points per domain name was 2 and it was seen that where the domain name had decided to implement TLS, the TLS was enabled across all of the MX points it operated. In terms of the percentage of servers within domains utilising TLS, the break down is shown Figure 5. The three domains identified in this graph .nz .ie .fi had low sample sizes that contribute to the high percentage with TLS enabled. Further work with a larger sample size may show different results.

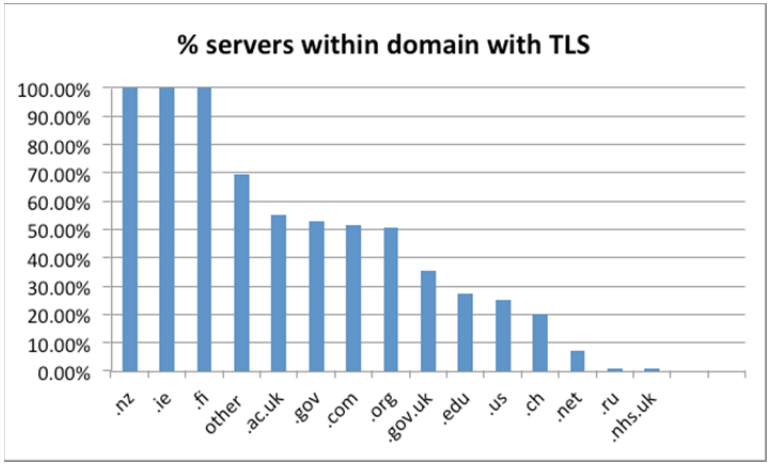

Figure 5: Percentage of Severs using TLS, by domain (Pilot Study)

On average, 37% of servers utilised TLS. It is quite clear from this sample that implementation of TLS is limited. The .ac.uk .gov and .org domains showed the highest implementations of TLS at around 50% of the sample with the other domains trailing off. What was is initially unexpected is the .nhs.uk domain space with 153 samples but only one with a TLS implementations. This is probably due to the addresses being harvested coming from web site general enquiry addresses with the majority of secure UK National Health Service (NHS) e-mail (NHSmail) traffic being routed internally via a separate private e-mail system that the NHS operate.

4.3. Larger-Scale Study: Results

Following on from the pilot sample, a larger survey of 17,148 MX points was undertaken using the same methods. It was discovered that 61% of domains surveyed had TLS enabled, compared with 54% in the pilot. The results show a higher overall percentage with TLS 47% of those surveyed by domain, compared to a figure of 37% in the pilot study; the results are illustrated in Figure 6.

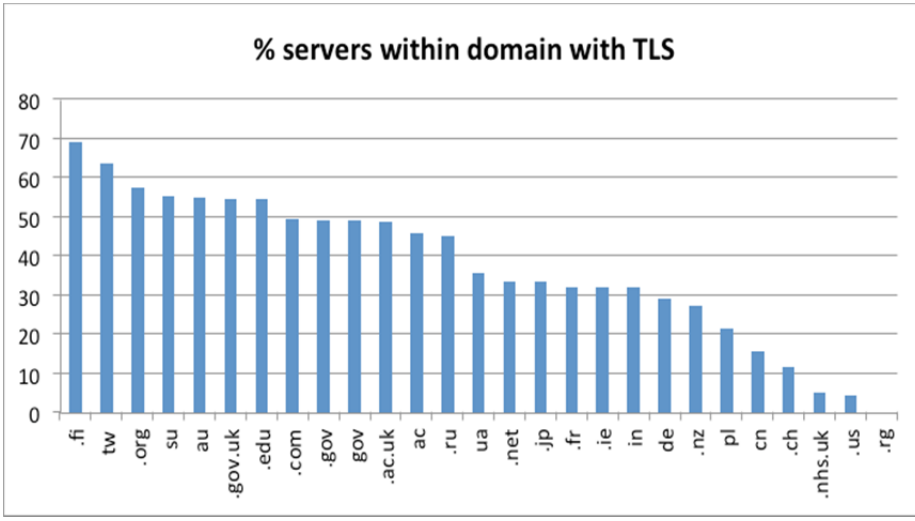

Figure 6: Percentage of Servers using TLS, by domain (Larger-Scale Study)

5. Discussion & Future Work

Our results provide a useful initial insight into the user perceptions of privacy in the online world and the implementation of technologies to support online privacy. Privacy cannot be guaranteed by a one-stop, single solution. Interventions range from policies and government legislation to technical solutions, such as encryption and protection from online attackers. As such, ensuring privacy is vast and coarse. Our future work focuses on developing a more detailed picture of the online privacy landscape by triangulating users experiences and perceptions with technical information, as exemplified in the work of this paper, to form taxonomy of recommended best practice for ensuring user privacy in today's online world.

To follow the initial user survey, a series of focus groups are planned. These are intended to gain an understanding of perceptions and practices related to individuals' use of computer systems in relation to issues of privacy.

Whilst a time consuming exercise, TLS surveying of a fuller sample of domains will provide a much more reliable analysis of the TLS implementation on the global Interet. In addition, it is intended that further work will look to address adoption of

other technical security measures that might ensure privacy, such as the usage of encryption, authentication and nonrepudication technologies.

6. References

Colombo, P.; Ferrari, E., "Towards a Modeling and Analysis Framework for Privacy-Aware Systems," Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom) , vol., no., pp.81,90, 3-5 Sept. 2012

Farrell, S. (2009). Why Don't We Encrypt Our Email?. IEEE Internet Computing Society , Volume 09 , pp. 1089-7801.

Goldberg, I., Wagner, D., & Brewer, E. (1997). *Privacy-enhancing technologies for the Internet*. CALIFORNIA UNIV BERKELEY.

Kemp, R. and Moore, A.D. (2007), "Privacy", Library Hi Tech, Vol. 25 Iss: 1, pp.58 – 78.

Kosa, T. A. (2010, August). Vampire Bats: Trust in Privacy. In Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on (pp. 96-102). IEEE.

Robison, C.; Ruoti, S.; van der Horst, T.W.; Seamons, K.E., "Private Facebook Chat," Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom) , vol., no., pp.451,460, 3-5 Sept. 2012

Solove, D.J. (2002), "Conceptualizing privacy", California Law Review, Vol. 90.

Warren. A. (2002), "Right to privacy? The protection of personal data in UK public organizations". New Library World, p. Volume: 103 Issue:11.

Weissa, S. (2009). Privacy threat model for data portability in social network applications. International Journal of Information Management, Volume 01, 2009.

Yee, G. (2006), "Measuring Privacy Protection in Web Services," *Web Services, 2006. ICWS '06. International Conference on* , vol., no., pp.647,654, 18-22, doi: 10.1109/ICWS.2006.87