Analysing security and privacy issues of using e-mail address as identity

Lei Jin, Hassan Takabi and James B.D. Joshi*

School of Information Sciences,
University of Pittsburgh,
410 IS Building, 135 N. Bellefield Avenue,
Pittsburgh, PA 15213, USA
Fax: (412) 624-2788
E-mail: lej17@pitt.edu
E-mail: hatakabi@sis.pitt.edu
E-mail: jjoshi@mail.sis.pitt.edu
*Corresponding author

Abstract: Nowadays, many websites allow or require users to use their e-mail addresses either as identity or for other purposes. Although username-based identity problems resulting from users' behaviours have been a research focus for quite some time, the serious issues of using e-mail address as identity and the associated online behaviours of users have not been well investigated. In this paper, we discuss and analyse security and privacy problems resulting from using e-mail address as identity via well-designed user behaviour survey and by investigating websites' design schemes. Our results illustrate that using e-mail address as identity poses high security and privacy risks. This is mainly because of the multiple usages of e-mail addresses and users' improper online habits. Moreover, we discuss drawbacks of existing solutions for e-mail address as identity and related password problems, and present potential solutions that may be used to secure online identity management systems in future.

Keywords: e-mail address identity; authentication; security.

Reference to this paper should be made as follows: Jin, L., Takabi, H. and Joshi, J.B.D. (2011) 'Analysing security and privacy issues of using e-mail address as identity', *Int. J. Information Privacy, Security and Integrity*, Vol. 1, No. 1, pp.34–58.

Biographical notes: Lei Jin is currently working toward his PhD at the School of Information Sciences, University of Pittsburgh and is a member of the Laboratory of Education and Research on Security Assured Information Systems (LERSAIS). He received his MSE in Software Engineering from Tsinghua University and his BS in Computer Software from Tsinghua University in 2009 and 2006 respectively. His research interests include authentication, privacy and security in social computing and in mobile computing, usable privacy and security. He is a student member of the IEEE and the ACM.

Hassan Takabi is currently working toward his PhD at the School of Information Sciences, University of Pittsburgh and is a member of the Laboratory of Education and Research on Security Assured Information Systems (LERSAIS). He received his MSc in Information Technology from Sharif University of Technology and BS in Computer Engineering (Software)

from AmirKabir University of Technology (Tehran Polytechnic) in 2007 and 2004, respectively. Before joining the University of Pittsburgh, he was a Research Scholar in the E-Security Research Centre at the London South Bank University. His research interests include access control models, trust management, privacy and web security, usable privacy and security, security, privacy, and trust issues in cloud computing environment. He is a student member of the IEEE and the ACM.

James B.D. Joshi is an Associate Professor and the Director of the Laboratory for Education and Research on Security Assured Information Systems (LERSAIS) in the School of Information Sciences at the University of Pittsburgh. He received his MS in Computer Science and his PhD in Computer Engineering from Purdue University in 1998 and 2003, respectively. His research interests include role-based access control, trust management, and secure interoperability. He is a member of the IEEE and the ACM.

1 Introduction

The rapid development of internet-based applications and services has made online websites a crucial part of our lives. More and more users are using online office services (e.g., Google Doc) instead of desktop office software, online social services (e.g., Facebook, Linkedin) to engage in social activities, online information boards (e.g., Twitter) to post their business products, and online shopping (e.g., Amazon) to do their purchasing. However, the security and privacy problems have also grown along with these newer and open Internet applications.

In order to use online services and applications, users typically need to create accounts including usernames and passwords. The username-based identity and the related password problems resulting from users' online behaviours have been a focus of research studies for quite some time (Florencio and Herley, 2007; Florêncio et al., 2007; Perlman and Kaufman, 2008; Just and Aspinall, 2009). Florencio et al. (2007) report the results of a large-scale study of password usage habits, where they discuss and analyse the password sharing rates among sites, password usage and strength. They also emphasise that it is better to increase the strength of usernames rather than passwords as it is more effective in defending against bulk guessing attacks than attacks that focus on compromising passwords (Florêncio et al., 2007). Perlman and Kaufman (2008) summarise the existing problems of username-based identity and password schemes, including the problems related to the use of the same usernames and passwords in multiple websites. Just and Aspinall (2009) investigate users' behaviours related to how they select or design secret questions and answers when they register accounts. Their results show that poorly-chosen questions and answers would weaken security of users' accounts.

However, serious issues related to using an e-mail address as an identity and related users' behaviours have not been well investigated. E-mail address as identity is becoming more common and may soon replace username-based identity as it has some advantages (Jin et al., 2010). First, it is easier for a user to remember his e-mail addresses rather than different usernames. A user may easily apply one of his e-mail

addresses as identity for multiple websites; hence, he only needs to remember one identity rather than multiple username-based identities. Second, web service providers find it more convenient to contact users using e-mail addresses that are collected during account registration phase. For example, e-mail communication as a confirmation and tracking method is important for online shopping websites, such as Amazon and EBay. Third, users are able to get or reset their passwords via e-mail when they forget them; this approach is regarded as convenient and safe by online service providers.

Because of the above-mentioned benefits, many websites allow and some even require users to use their e-mail addresses as identity in their login pages. However, using e-mail address as identity introduces severe security and privacy risks. It can potentially expose more private information to the public than a username-based identity, thus introducing significant privacy concerns. Such concerns arise because an e-mail address identity, in addition to representing a unique ID, reveals the user's e-mail address details (e.g., login name, domain name, etc.). For instance, if a user has an e-mail address identity *mary@gmail.com* in one website, it reveals the fact that she has an account named *mary* in the Gmail service. Therefore, once a user's e-mail addresses or e-mail address identities have been compromised, it may put his other related accounts at a higher risk.

In this paper, we discuss and analyse security and privacy issues related to using an e-mail address as an identity. We illustrate how serious they are, identify potential attacks to e-mail address identity systems and analyse the extent of damages such attacks may cause. We present results of our survey which was carried out to investigate users' behaviours in using an e-mail address as an identity. We also investigate popular websites' account design schemes to see how they address security and privacy issues related to users' e-mail addresses. Based on our user survey and investigation of online sites, we show that security and privacy problems of using e-mail address as identity are not caused only by user behaviours but also by design schemes used by websites. We demonstrate that the use of an e-mail address as an identity poses a higher risk to users' accounts and may result in more security breaches than username-based identity. This is mainly due to multiple usages of e-mail addresses, such as identification and communication. We also argue that using an e-mail address as an identity cannot be a secure solution for users because users share their identities and passwords in multiple websites – such sharing behaviour can put all of a user's accounts at a higher risk when any one of his e-mail accounts used as identity is compromised. Furthermore, we present the drawbacks of existing solutions for e-mail address identity (and the related password problems), and propose suggestions for users to better protect their e-mail address identities. We also discuss other potential solutions that help secure identity management systems.

The rest of the paper is organised as follows. Section 2 introduces our methodology for the user behaviour survey and website investigation. In Sections 3 and 4, we analyse the results of our user behaviour survey and present the investigation of websites, respectively. In Section 5, we discuss existing solutions for the identified security and privacy problems, their drawbacks and introduce possible approaches. Finally, in Section 6, we present our conclusions.

2 Methodology

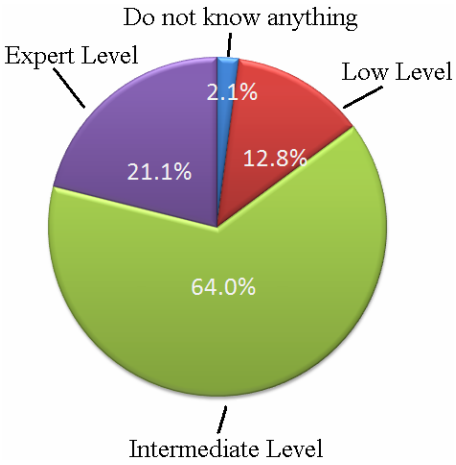
2.1 The survey on the use of e-mail addresses

Our survey is focused on investigating users' behaviours related to the use of their e-mail addresses. The survey includes 431 participants in total, 357 of them responded to the survey through Amazon Mechanical Turk (Amazon, 2010) and 74 of them responded to the survey via e-mail.

There are 244 male and 187 female survey participants. Majority of the participants' (271) ages are between 20 and 30, while 112 of them are middle-aged individuals between ages 30 and 50. There are 33 and 15 participants who are younger than 20 and older than 50, respectively. With regards to educational background, 125 participants have the undergraduate degree whereas 270 participants have the graduate degree or above. Only 36 participants do not have any college degree.

The participants were also asked about their information security background. As shown in Figure 1, most of the participants (64.0%) claim that they have an intermediate knowledge level of information security while 21.1% of the participants believe that they are experts in information security. Only 12.8% and 2.1% of the participants said they know little and nothing about information security, respectively.

Figure 1 Users' assessments of their knowledge of information security (see online version for colours)



In order to collect information about users' behaviours in using e-mail addresses, we designed a questionnaire and asked each participant to carefully read the questions and answer them. The questionnaire included 31 questions about using e-mail addresses in general, using e-mail addresses when registering in different websites, using the same passwords for different e-mail address identities, etc. The questions were in the form of multiple choice and yes/no questions. Particularly, we were interested in the following user's behavioural information:

- *Use of e-mail addresses:* We focus on the questions that include the following: How many e-mail addresses each participant has? How many participants' e-mail addresses, divided by primary and non-primary, are used during website registrations? Do the participants use the same passwords for their e-mail accounts and e-mail address identities in other websites?
- *Username preferences for e-mail accounts:* We gather information of participants' preferences with regards to using the same usernames for their e-mail accounts.
- *Capability of remembering passwords for e-mail address identities:* We examine the pairs of e-mail accounts/identities and passwords participants can remember.
- *Users' attitudes for securing e-mail address identities:* We collect their attitudes for several popular password schemes and possible solutions in future.

We collected quantitative data through the survey and used it to analyse the relationships between the users' behaviours and potential security and privacy problems related to these behaviours. In order to analyse the data, we designed a table that includes a code for each question and the corresponding number of possible answers.

2.2 Website investigation

In addition to the users' behaviours, another factor that may cause security and privacy issues when using online services and applications is how websites are designed. To this end, we have investigated popular websites to examine security and privacy risks caused by account design schemes used by these websites. We have tried to infer how users' private information is stored in these websites. Based on the inferences, we have demonstrated that users' accounts in different websites have a high risk of being compromised when their e-mail accounts or e-mail address identities in some sites are compromised and passwords are exposed, because users may have used the same e-mail address identities and passwords in multiple sites.

In the study of websites, we examined 120 popular websites that can be categorised into different types of online services, such as shopping websites (e.g., Amazon and Newegg), social networking websites (e.g., Facebook and Linkedin), financial websites (e.g., Paypal), etc. In our study, we specifically examined the registration, login and password recovery processes. We collected information related to the following three components of each website:

- *website's registration scheme:* we focus on what types of users' information is requested during registration
- *types of identities:* we look at what kinds of identities are supported by the websites
- *username and password transmission schemes:* we investigate whether usernames and passwords are transmitted as plaintext to the web services
- *password recovery schemes:* we examine what types of password recovery schemes are applied by the websites.

The questionnaire, user behaviour statistic reports and analysis report of websites are available in Jin (2010).

3 Security and privacy issues of e-mail address usage

In this section, we analyse the results of our user behaviour survey and discuss various security and privacy problems related to the use of an e-mail address as an identity.

3.1 E-mail address usage

3.1.1 Number of e-mail accounts and popular e-mail services

The survey results show that most of the participants (95.8%) adopt e-mail as their main medium to communicate with others. 69.6%of them have between one and three mail accounts while the remaining have more than four e-mail accounts. However, 25% of the participants use only one e-mail account frequently, 42.2% of the participants often use two e-mail accounts and 23.6% of them continually use three e-mail accounts. In summary, approximately 90% of the participants use up to three e-mail addresses frequently although they may have registered more e-mail accounts.

The survey also shows that 86.4% of the participants use Gmail as one of their primary e-mail accounts; 14.8% and 46.5% of them use Hotmail and Yahoo Mail, respectively, as one of their primary e-mail accounts. And, 23.9% of the participants use other e-mail addresses, such as their company e-mail addresses as their primary e-mail accounts.

3.1.2 Primary e-mail addresses

As indicated in Figure 2, 71.8% of the participants use their primary e-mail addresses while registering in other websites. These e-mail addresses are used by online services as identities to contact users for various purposes such as notifications, shopping confirmations and password recovery. Figure 3 illustrates the distribution of websites where primary e-mail addresses of the participants are used. 56.1% of the participants use their primary e-mail addresses on shopping websites (e.g., Amazon, Ebay and Newegg); 36.4% of them employ their primary e-mail addresses on financial websites (e.g., Paypal and Citibank); 55.1% of them use their primary e-mail addresses on social networking websites (e.g., Facebook and Twitters) and 8.1% of them adopt their primary e-mail addresses in other websites.

Figure 2 Use of primary e-mail address (see online version for colours)

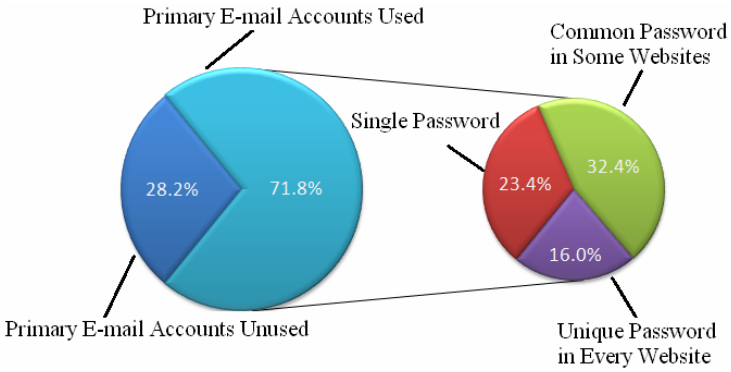
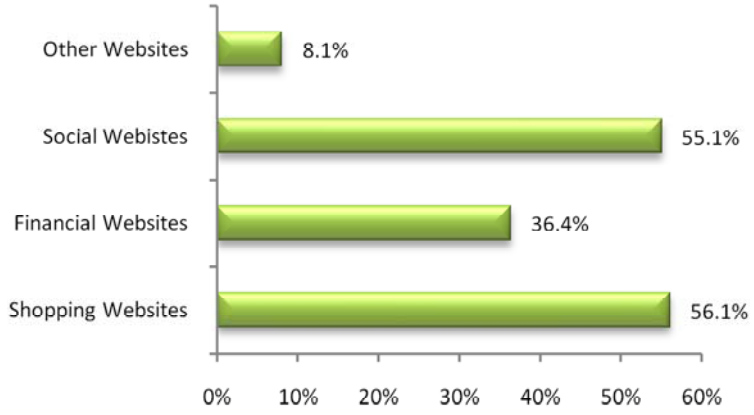


Figure 3 Distribution of primary e-mail addresses used in different types of online websites (see online version for colours)



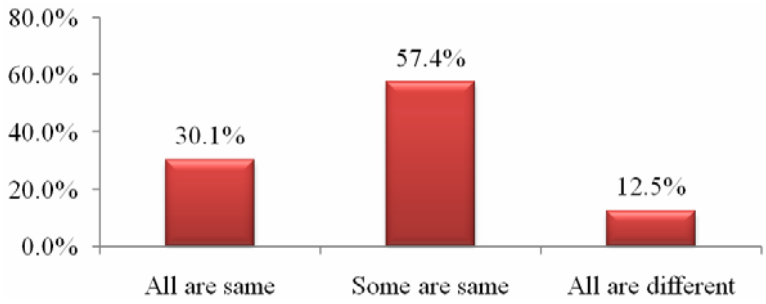
The survey also shows the number of the primary e-mail addresses used by participants during website registration. 47.6% of them use one of their primary e-mail accounts for website registrations; 41.1% of them use two or three of their e-mail addresses and approximately 10% of them apply all of their primary e-mail accounts during website registrations.

Florencio and Herley (2007) have shown that people usually use the same passwords on average in three different websites. In our survey, we also investigate how many participants use their e-mail accounts' passwords on other websites where their e-mail addresses are used as identities or for other purposes. As shown in Figure 2, of the 71.8% of the participants using their primary e-mail addresses during registration, 55.8% (23.4% + 32.4%) of the participants share at least one of their primary e-mail accounts' passwords with the e-mail address identities on other websites. In particular, 23.4% of the participants use one password for their primary e-mail accounts and e-mail address identities on multiple websites. Only 16.0% of all the participants said they use their primary addresses for website registration processes, but never use the same passwords as that of their primary addresses for the e-mail address identities in multiple websites.

3.1.3 Non-primary e-mail addresses

Although 28.2% of the participants (see Figure 2) never use their primary e-mail accounts on other websites, all of these participants employ at least one of their non-primary e-mail addresses for website registrations; this may be because some websites, such as Amazon and Facebook, only support e-mail address identity. Compared to the impact of sharing the same e-mail passwords between primary e-mail addresses and corresponding e-mail address identities, our survey results (see Figure 4) demonstrate that participants are at a higher risk in using their non-primary e-mail addresses than that of primary e-mail addresses. A significant proportion of the participants either share fully (30.1%) or partially (57.4%) their non-primary e-mail accounts' passwords with the corresponding e-mail address identities on other websites. Only12.5% of the participants indicate that they never use their non-primary e-mail accounts' passwords for the corresponding e-mail identities on other websites.

Figure 4 Passwords sharing rate among non-primary e-mail addresses (see online version for colours)



3.1.4 E-mail account settings

We also collected participants' behavioural information regarding how they configure settings in their e-mail accounts. 51.1% of the participants connect at least one of their primary e-mail addresses with their non-primary e-mail addresses for some purpose, perhaps to receive e-mails from other e-mail accounts and/or to receive password reset links when they forget passwords of their e-mail accounts. In particular, 43.0% of the participants set their primary e-mail accounts to receive e-mails from their non-primary accounts. So, we can infer that:

1 the participants' e-mail accounts may be easily compromised when their other e-mail accounts, which are used to receive password reset links of these accounts, are compromised

2 the participants may expose their non-primary e-mail accounts more when their primary e-mail accounts are compromised

3 the participants may lose their accounts on other websites once their e-mail accounts used on these websites are compromised.

This is because the e-mail containing plaintext passwords or password recovery links of these accounts can be assumed to be exposed. For instance, assume one user allows his Gmail account to receive his e-mails of Yahoo Mail and he uses his account of Yahoo Mail as the identity on Amazon. Once this Gmail account is compromised, his Amazon account will likely be compromised as well because an adversary can get the e-mail containing Amazon's password reset link (it has been originally sent to his Yahoo Mail account by Amazon) in his Gmail accounts.

3.1.5 Discussion

Based on the results related to use of primary e-mail addresses, we can infer that the percentage of participants who use the same passwords in both their primary e-mail accounts and websites where these e-mail addresses are used as identities, is 71.8% × 23.4% = 16.8%, where 71.8% is the portion of the participants using their primary e-mail addresses on other websites, and 23.4% is the portion of the participants using the same password both for their primary e-mail accounts and for accounts of websites where these e-mail addresses are used as identities.

Combining the results-related to the popular e-mail services, we can infer that the percentage of participants who use their primary e-mail addresses as identities (e.g., *Gmail*) on the social websites is approximately 86.4% × 55.1% = 47.6%, where 86.4% is a portion of the participants using Gmail as one of their primary e-mail addresses and 55.1% is the portion of the participants employing their primary e-mail addresses as identities on social websites. Assuming that 30% of the participants use their Gmail accounts in Facebook, the portion of the participants having the same passwords for their Gmail and Facebook accounts is approximately 16.8% × 47.6% × 30% = 2.4%. We believe that these participants will have both accounts compromised if their Gmail accounts or Facebook accounts are compromised and then the passwords are exposed.

Through the analysis above, we can infer that adversaries only need to compromise a user's primary account, e.g., his Gmail account, in order to potentially compromise many of his other accounts on multiple websites. For example, an adversary may compromise a user's other e-mail addresses/identities on multiple websites by trying the same password used in his Gmail account, checking Gmail account settings and searching registration confirmation and password reset e-mails. Moreover, the adversary can aim to compromise a user's account on some poorly-designed website where the user employs his Gmail address as identity and share the same password as that of this account with his Gmail account. After the password of this account is exposed to the adversaries, they can compromise his Gmail account and further compromise his other accounts on multiple websites. We analyse the issues related to the website design in Section 4.

3.2 Username diversity

We also investigate the issue of participants' preferences regarding usernames in different e-mail services. The results show that 41.1% of the participants indicate each using a single username for all of his e-mail accounts in different e-mail services while58.9% of the participants indicate each applying one username for several e-mail services but not for all of the e-mail services for which they are registered. From the latter group, 37.2% of the participants indicate each using two different usernames, 38.1% of the participants indicate each applying three different usernames and 24.7% of the participants indicate each adopting more than three usernames for his e-mail accounts. In addition, we found that more than 60% of the participants have special preferences for their e-mail accounts' names. This result indicates that users are inclined to use as few usernames as possible because of some understandable reasons, such as the effort to reduce the difficulty in remembering all the usernames.

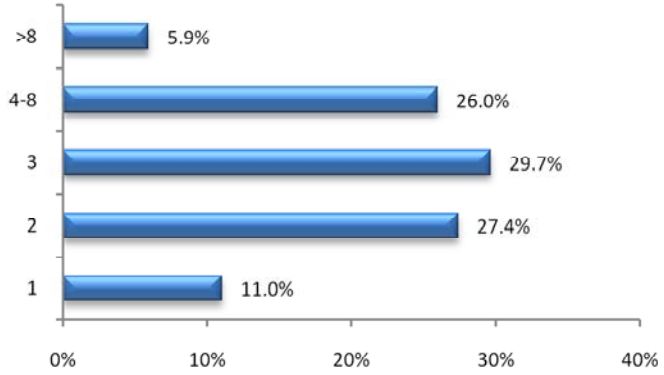
The investigation of username diversity in e-mail accounts reveals that it is not difficult to obtain all of the e-mail addresses that belong to one user, because most of the users only use between one and three usernames for different e-mail services. This user behaviour also shows that an adversary can easily guess users' e-mail identities on different websites.

3.3 Capability to remember passwords

We also asked the participants how many pairs of e-mail accounts/identities and corresponding passwords they are able to remember. As shown in Figure 5, more than half of the participants (27.4% + 29.7% = 57.1%) indicate that they can remember two to three pairs of e-mail addresses/identities and passwords while 26.0% of the participants

said that they can remember four to eight passwords. Only11.0% of the participants indicate that they can remember one pair, while 5.9% of them indicate a higher capability to memorise and claim they can remember more than eight pairs.

Figure 5 Capabilities to remember e-mail accounts/identities and passwords (see online version for colours)



Compared to the study on capability to remember with regards to passwords in Florencio and Herley's results (2007) where the authors show that their participants can usually remember five or six different passwords for username-based identities, our results indicate that the participants may have less memory retention capabilities for e-mail addresses/identities and passwords. Therefore, we may infer that some of the participants may be inclined to use the same passwords as that of their e-mail accounts for their e-mail address identities on multiple websites. This may validate our previous analysis that using the same passwords among users' e-mail accounts and corresponding e-mail address identities on multiple websites is a common practice.

3.4 Comparison: male vs. female

3.4.1 Confidence in security background

The Male participants are more confident in their security background: nearly 30% of the male participants claim that they have an expert level. However, only 8.2% of the female claim that.

3.4.2 Use of primary e-mail

Majority of participants, male (68.1%) or female (73.3%) use their primary e-mail addresses during website registration. However, the result suggests that 30.0% of male participants like to share one password between their primary e-mail accounts and websites where their primary e-mail addresses are used, compared to 14.7% of female participants.

Another finding in our survey is that female participants are more likely to use as identities their primary e-mail addresses in shopping sites (60.2%) and social sites (58.3%) compared to the male participants (54.1% for shopping sites and 53.6% for social sites). However, male participants use their primary e-mail accounts for financial sites (38.6%) more than female counterparts (26.7%).

3.4.3 Non-primary e-mail

In the use of non-primary e-mail accounts, we get the similar results as that of the use of primary e-mail accounts with regards to the issues of sharing passwords between the male and female participants. Compared to the portion of the female participants (18.7%), more male participants (34.9%) indicate each sharing one password between his primary e-mail accounts and the websites where his primary e-mail addresses are used.

3.4.4 E-mail account setting

The survey shows that 52.7% of the male participants connect at least one of their primary e-mail addresses with their non-primary e-mail addresses while only 42.0% of the female participants apply such same settings. While 45.9% of the male participants use their primary e-mail accounts to receive e-mails from their non-primary accounts, only 30.7% of the female participants do that. Based on this result, we infer that the male participants' e-mail addresses are more dependent on each other and they may lose more e-mails or e-mail accounts once one of their e-mail accounts is compromised.

3.4.5 Username diversity

Our result shows that 43.4% of the male participants indicate each having only one username for their e-mail addresses, while this percentage for the female participants is 37.8%. For the participants who use at least two usernames, most of the male and the female participants adopt two or three different usernames.

3.4.6 Capability to remember passwords

The percentage of the male participants (39.2%), who remember one or two pairs of usernames and passwords, is more than that of the female participants (35.3%). However, the portion of the female participant (64.7%), who can remember more than three pairs of usernames and passwords are more than that of the male group (60.8%). This result shows that the female group has the higher retention capability than the male group.

3.4.7 Summary

Based on the comparisons discussed above, we can infer that the female group is more careful about their activities that may cause security problems. The male group is less cautious, although the result of self-assessments for their security background shows that they may know more about security issues than the female group does.

3.5 Comparison: IT vs. non-IT groups

3.5.1 E-mail service adoption

An interesting finding regarding e-mail selection between individuals in IT group (people with IT as educational background) and general people (i.e., non-IT people) is that IT persons prefer to use Gmail (82.4%) and the e-mail services that directly connect with their work or study, such as the company and university e-mail services(55.4%). They

may not like to use Yahoo e-mail (5.4%) accounts. However, general individuals prefer to use Yahoo e-mail service (54.5%) while they may not use their e-mail addresses at work (17.1%).

3.5.2 Primary e-mail

There are no big differences in the numbers of primary e-mail accounts between individuals in IT group and non-IT group. Most of them usually have one to three primary e-mail addresses. However, general people (72.8%) are more likely to use their primary e-mail addresses during website registrations, compared to IT people (67.6%). More people in non-IT group (35.5%) prefer to share the same password between their primary e-mail accounts and the websites where their primary e-mail addresses are used. The proportion in IT group is 20%.

Based on above comparisons, we argue that non-IT individuals use fewer primary e-mail addresses and their primary e-mail addresses may be easier to compromise due to their behaviour with regards to sharing passwords.

3.5.3 Non-primary e-mail

In the use of non-primary e-mail addresses, the proportion of the participants who share one password among e-mail accounts and the websites where those e-mail accounts are used as identities, is not significantly different between the users in IT group (29.7%) and general users (30.2%). However, the percentage of users who partially share their passwords among e-mail accounts and e-mail address identities is different – 61.7% of the participants in general group and only 46% of the users in IT group. This result also indicates that fewer participants in general group do not share passwords of their e-mail accounts with their corresponding e-mail address identities.

3.5.4 E-mail account setting

In IT group, 56.8% of the participants indicate each connecting at least one of his primary e-mail addresses with his other e-mail addresses for some purpose, while that proportion in the non-IT group is 49.9%.

With regards to the issue of using the primary e-mail accounts to receive e-mails from their non-primary accounts, 52.7% of the participants in IT group demonstrate such behaviour, while it is only 40.9% for the participants in general group.

These two comparisons reflect that IT persons may be more familiar with using e-mail addresses. However, it also indicates that IT professionals may lose more e-mail accounts once one or some of their e-mail accounts have been compromised.

3.5.5 Username diversity

35.1% of the participants in IT group use one username for all of their e-mail accounts in different e-mail services, while 42.8% of the participants in general group have the same behaviour. However, the portions of the participants, who adopt two to three different usernames for their e-mail accounts, occupy major parts and are nearly same between the IT group and the general group.

3.5.6 Capability to remember passwords

As shown in Figure 6 and Figure 7, in general, individuals in IT group can remember more pairs of username and password. Specifically, fewer individuals in IT group only remember one (8.1%) or two (16.2%) pairs of username and password, compared to those in general group (11.8% and 30.4%, respectively). More than 40% of the users in IT group can remember three pairs of username and password while only 26.4% of the participants in general can remember three pairs. In addition, 33.8% of the participants in the IT group can remember four to eight pairs of username and password while only 23.9% can do that. However, there are 7.5% of the non-IT people who said that they remember more than eight pairs while no participant in IT group has such claim.

Figure 6 IT group: memory retention capability (see online version for colours)

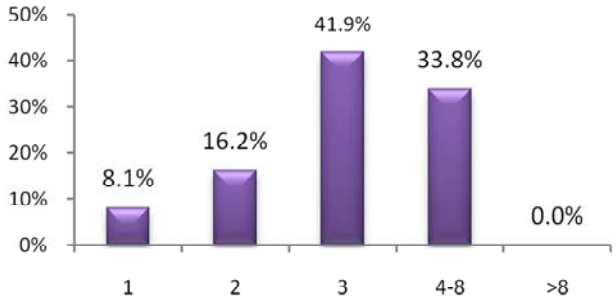
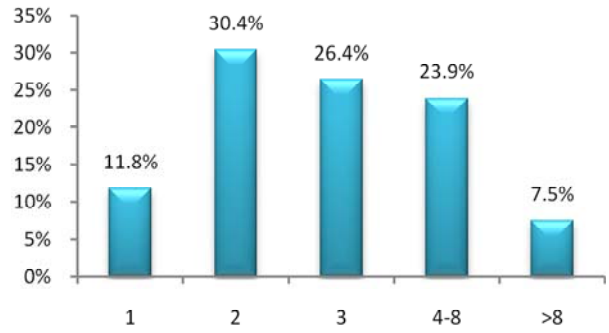


Figure 7 General group: memory retention capability (see online version for colours)



3.5.7 Summary

From the above comparisons between the IT and No-IT groups, we confirm the argument that the general population might be less aware/educated about security issues related to the use of e-mail address.

3.6 E-mail address as privacy risk

Based on our previous discussion and analysis, we can conclude that e-mail addresses are a huge privacy risk. This is also confirmed by the findings of our survey where 72.3% of the participants regard their e-mail addresses as their private information and 73.7% of

the participants believe that exposure of e-mail addresses is a privacy violation. In particular, 65.4% of the participants are sensitive about two issues:

1 e-mails can disclose the locations of the senders

2 internet service providers (ISP) can store user's e-mails in their servers.

However, it seems difficult to conceal users' e-mail addresses and handle above issues. The difficulties come from the following:

- users have to provide their e-mail addresses when they send e-mails to others
- users may share all the e-mail addresses of a group to everyone in the group when they send e-mails
- design of websites encourages or forces users to share their e-mail addresses
- it may be difficult to make the international law for protecting the locations of the e-mail senders and the content of the e-mails.

We will discuss the existing and potential solutions to handle above issues in Section 5.

3.7 Survey summary

Based on the analysis and discussion in this section, we argue that participants may not act as their claims, compared to their self-assessments for the security knowledge in use of e-mail accounts. No matter whether the individuals are in the IT or the non-IT group, their incautious behaviours put their e-mail accounts and e-mail address identities at a higher risk.

One of the questions in the survey supports the above argument. Secure sockets layer (SSL) is cryptographic protocol that provides communications security over the internet. Our survey indicates that 74.7% of the participants in general population are not aware of SSL or do not log into websites using SSL enhanced pages, although most of the participants in IT group usually use secured login pages.

4 Security and privacy issues related to website designs

Security and privacy problems of using e-mail address as identity arise not only from users' habits but also from websites' lack of secure designs and mechanisms. In this section, we discuss these issues and illustrate that additional risks are created when poorly-designed websites are compromised, exposing the private information stored in these websites.

4.1 Design schemes

In this section, we discuss our findings based on the investigation of 120 popular websites with regards to their schemes related to account registration, types of identities and password recovery. The details of the investigation are shown in the webpages indicated in Jin (2010).

The websites in our investigation can be categorised into the following types: shopping, social network, living and e-mail services. In detail, we have analysed 25 shopping sites (such as Amazon, Ebay and Newegg), 77 social network sites (such as Facebook, MySpaces and LinkedIn), 14 sites related to living (such as Expedia, Craigslist and Yellow Page) and four e-mail services (Gmail, Yahoo Mail, Hotmail and AoL Mail).

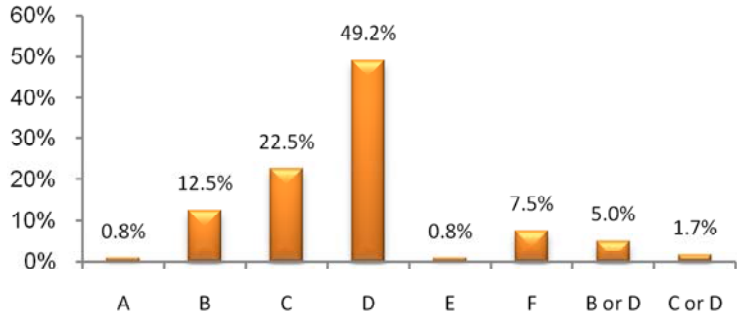
4.1.1 Website registration schemes

Our findings of the registration schemes that websites generally adopt (Table 1) are illustrated in Figure 8. The proportion of the websites that request users to input e-mail addresses as identities or use e-mail addresses for other purposes (e.g., for password recovery) is 81.7% (Scheme C + Scheme D + Scheme E + Scheme F + Scheme C or Scheme D); 20.8% of the websites (Scheme B +Scheme E + Scheme F) request users to select or define their secret questions and answers. In particular, 5.0% of the websites support two kinds of registration schemes (Scheme B and Scheme D): when users forget their passwords, they can get or reset passwords via their e-mail addresses or by answering their secret questions. There is also 1.7% of the websites (scheme C and Scheme D) where users can choose an e-mail account or the combination of usernames and e-mail accounts as identities. Only 13.3% of the websites (Scheme A + Scheme B) do not request users' e-mail addresses when registering.

Table 1 Catalogue of website registration schemes

Scheme	*Description*
A	Websites require users to input unique usernames.
B	Websites require users to input unique usernames and select or customise secret questions and answers.
C	Websites only require users' unique e-mail addresses.
D	Websites require users to input unique usernames and unique e-mail addresses.
E	Websites require users to select questions and set the corresponding answers, or ask users to define customised questions and provide the answers. The e-mail addresses are also required by the websites.
F	Websites require users to input unique usernames, unique e-mail addresses, secret questions and the corresponding answers.

Figure 8 Distribution of website registration schemes (see online version for colours)



In particular, we found that all the shopping sites request users to input their e-mail addresses when they register, while 12 out of 14 sites in the living category and 76.6% of the social sites require users' e-mail addresses.

The results above confirm our argument that most of the websites request users to input their e-mail addresses for some purpose, such as to send passwords or password reset links. This fact also reveals a high risk for e-mail accounts because users utilise their e-mail addresses on many different websites.

4.1.2 Types of identities

We also investigate the types of identities that websites support. The result shows that 55.8% of the websites only support e-mail address as identity, 27.5% websites employ username-based identity and 16.7% of the websites support both of these identities. Based on this data, we find that e-mail address identity is widely adopted: 72.5% (55.8% + 16.7%) of the websites support it.

Our further analysis shows that more than 60% of shopping sites, living category sites only adopt e-mail address as identity, while nearly 50% of social websites have such activity. In addition, we find only 1 in 14 living sites allow users to login either by their usernames or their e-mail addresses. However, these portions in shopping and social sites are more than 22%. Based on this result, we argue that shopping and social websites better convenience to users with regards to the login process.

Based on the data in website registration schemes, we can infer that the proportion of websites that request users' e-mail addresses as identities is approximately81.7% × 72.5% = 59.2%, where 81.7% is the percentage of the websites which require users' e-mail addresses during registration and 72.5% is the portion of the websites that adopt e-mail addresses as identities. This indicates that e-mail address as identity is becoming popular. However, our analysis and discussion in Section 3 show that e-mail address as identity poses huge security and privacy risks.

4.1.3 Username and password transmission schemes

Besides the investigation for website registration schemes and types of identities, we investigated websites' transmission mechanisms for users' usernames and passwords using a sniffer tool (EffeTech, 2009) as well.

The results of the examinations show that 74.2% of the websites encrypt username and password transmission, while other 25.8% adopt unsecure username and password transmission schemes (transmit the data as plaintext). In particular, 18.3% of the websites do not have a secure transmission scheme. When users log into one of those unsecure websites, the adversary can compromise their accounts by using sniffing tools to obtain their usernames and passwords.

Our result shows that all the e-mail services use secure mechanisms for username and password transmissions while 92.0% of the shopping sites and 85.7% of the living sites are secure in username and password transmission. However, only 64.9% of the social sites have such support. We believe that social sites should pay more attentions to securing data transmission since they are becoming more popular and users may share more sensitive information in them.

4.1.4 Password recovery schemes

Based on the password recovery schemes of the websites that we studied (Table 2), our collected data is shown in Figure 9. When users forget their password, 58.3% of websites (Scheme B1) send a password reset links to users' e-mail addresses and 8.4% of them (Scheme C1 + Scheme C2) ask users to answer their customised secret questions. Of the remaining websites, 14.2% of them (Scheme A1) send users' original passwords as plaintext and 7.5% of them (Scheme A2) send new passwords as plaintext to users' e-mail addresses. Also, a small set of websites (Scheme A3) send temporary links displaying new passwords, or links showing secret questions to users, or send password reset links after users can answer secret questions correctly. In particular, 5.8% of the websites support two kinds of password recovery schemes: sending password reset links to users via e-mail (Scheme B1) and asking users to answer their pre-defined secret questions (Scheme C1), while 1.7% of them support the other two kinds of password recovery schemes: sending password original passwords in plaintext (Scheme A1) and sending password reset links to users (Scheme B1) via e-mail.

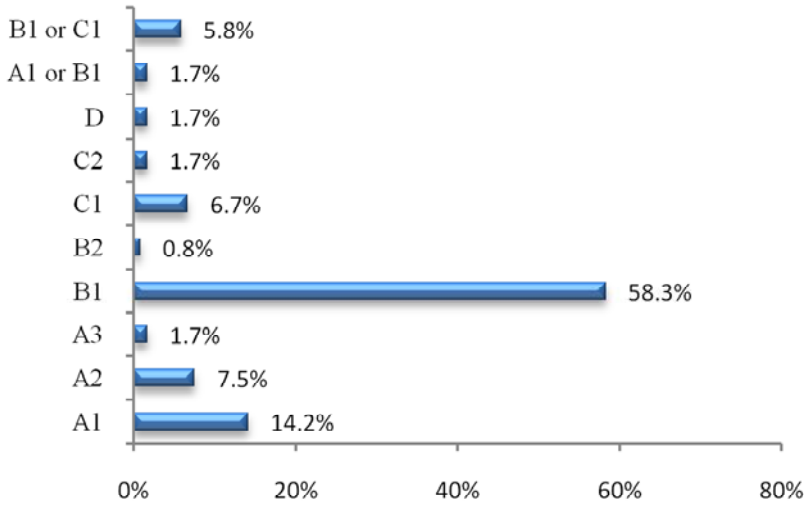
Table 2 Catalogue of password recovery schemes

Scheme	*Description*
A1	Users input their usernames or e-mail addresses. Services send the original passwords in plain text to the e-mail addresses recorded.
A2	Users input their usernames or e-mail addresses. Services send the new passwords set by services in plain text to e-mail addresses recorded.
A3	Users input their usernames or e-mail addresses. Services send temporary passwords (valid for a limited time) to users' e-mail addresses in plaintext. After users login using these temporary passwords, they can modify their passwords.
B1	Users input their usernames or e-mail addresses. Services send password reset links (valid for a limited time) to users' e-mail addresses.
B2	Users input their usernames or e-mail addresses. Services send the links (valid for a limited time) of pages displaying new passwords to users' e-mail addresses.
C1	Users input their usernames or e-mail addresses. Services provide users' selected or customised secret questions. After answering correctly, users can reset their passwords.
C2	Users input their usernames or e-mail addresses. Services provide users' selected or customised secret questions. After users answer the secret questions correctly, services send password reset links (valid for a limited time) to users' e-mail addresses.
D	Users input their usernames or e-mail addresses. Services send links that show users' selected or customised secret questions to users' e-mail addresses. After answering correctly, users can reset their passwords.

We also observe that more than 90% of the websites (Scheme A and Scheme B) send plaintext passwords or password reset links to users' e-mail accounts. In particular, 76% of the shopping sites, 88.3% of the social sites and 92.9% of the living sites apply such password recovery schemes. For the e-mail services, every site support the scheme that is to send password reset links to users' e-mail addresses. This confirms the high risk related to e-mail address identity discussed in the previous section: a user's accounts on different websites may be easier to compromise once some of his e-mail accounts have

been compromised. This is because adversaries can acquire or reset victim's passwords for his other accounts on multiple websites via the compromised e-mail account.

Figure 9 Password recovery schemes (see online version for colours)



Some users may believe that recovering passwords via secret questions is a secure mechanism. However, we argue that this is still not safe because users may typically apply the same secret questions and answers for multiple websites. Moreover, adversaries may obtain users' passwords through other means even though they cannot compromise their e-mail accounts in secure e-mail services. To do this, they can compromise poorly-designed websites and get the private data, such as users' e-mail addresses, passwords, secret questions and answers. After that, they may compromise accounts for other websites where users share e-mail address identities, passwords, or secret questions and answers. In the following, we analyse and discuss these issues.

4.2 What adversaries obtain after compromising websites

Generally, users' private data, including passwords and answers to secret questions, are stored on the websites. When adversaries have compromised the websites, they may be able to obtain these pieces of information. In the following, we discuss these issues based on our study of websites.

4.2.1 Password storage

We can infer that some websites store users' passwords in plaintext because they send users' original password in plaintext when the passwords are forgotten. These websites comprise at least 14.9% (Scheme A1 in the Figure 9) in our study, such as Pizza Hut (pizzahut.com) and Conduit (conduit.com). Although this is a small portion of the websites studied, we cannot ignore them because adversaries may obtain users' passwords easily after they compromise these websites.

In other websites, they may store users' passwords as hashed values in order to protect users' privacy. However, it is not difficult for adversaries to compromise these hashed passwords if the salt mechanisms (Dierksand and Allen, 1999), which can complicate and defend against dictionary attacks effectively, are not applied. This is mainly because users usually apply simple passwords which cannot be defended against a brute-force or dictionary attack because of the increased computing resources available currently. Florencio and Herley (2007) report that most of the users define their passwords to be 6–13 characters long. They report that more than 60% of such long passwords are composed only of lowercase letters; at least 10% of the 7–13 character long passwords are composed only of digits; only 10% of the passwords are composed of four different types of characters including lowercase/uppercase letters, digits and symbols. Based on these data, we can determine the number of calculations necessary to compromise users' hashed passwords. For example, in order to decrypt the eight-character long passwords that are composed only of lowercase letters or only of digits, the adversary only needs to try $(26)^8 + (10)^8 = 2.1 \times 10^{11} + 10^8 \approx 2.1 \times 10^{11}$ different passwords. It is not difficult for an adversary to search and compare the hashed values in this space to compromise approximately 70% (60% + 10%) of all eight-character long passwords.

As a result, hashed-password systems without salt mechanisms can no longer be considered immune to attacks. Adversaries can first target the websites that are easier to compromise. After getting users' hashed passwords, they can compromise them by the method discussed above. Finally, they can try these real passwords in different websites to compromise users' other accounts.

4.2.2 Secret questions and answers

Some well-designed websites (about 8%, Scheme C in the Figure 9) ask users to select or customise secret questions and answers. However, we strongly believe that users also employ their secret questions and answers on multiple websites. One reason for this behaviour may be the limited memory retention capability of users.

For the secret questions, most of these websites (except Scheme D) expose them to public. For the answers to the secret questions, some of the websites may store them as hashed values while some may store them in plaintext. After an adversary compromises a website, he may obtain the user's secret questions and the answers (if the answers consist of simple characters that are hashed without salts, adversaries may compromise them via the brute-force approach, as described above). Then, adversaries can check and compare secret questions on multiple websites, and try the answers they get to compromise more accounts. Even though adversaries cannot derive the text from the hashed answers, they can find or infer some of the users' private information from the data exposed on multiple social websites (Zheleva and Lise, 2009) to guess the answers to the secret questions.

5 Existing solutions and their drawbacks

In this section, we discuss several existing solutions to deal with security and privacy issues related to using an e-mail address as an identity and argue that these approaches are still far from providing strong protection. We only consider the identity mechanisms

which can be applied by websites that are open, dynamic and have a large number of users. The solutions for identity management systems based on cryptography may not be suitable for these websites because these mechanisms are suitable typically only for a controlled environment with well-organised key management schemes in place (DeadMan's Handle Ltd., 2005).

5.1 Password solutions

5.1.1 Password schemes

Users may believe that complex textual passwords can protect users' accounts effectively. Although this might be true, the fact is that this argument overlooks the usability issues. Complex passwords, composed of lowercase letters, uppercase letters, digits and symbols, are difficult for users to remember and are not easy to input when users log into the websites. Furthermore, even though some websites require users to set more complex passwords, these passwords will likely be reused by users in other websites. Therefore, the complex password scheme cannot solve the security problems of password usage we discussed earlier.

Some researchers believe that graphical passwords (Everitt et al., 2009) or biometric passwords (James, 2008) can significantly improve the security of users' accounts as these password schemes increase the length and complexity of passwords and may be more convenient to remember. With graphical passwords, a user only needs to remember a special image as a password, while biometric passwords do not require users to remember anything at all. However, we argue that these password schemes may not be widely adopted by websites. For graphical passwords, users may have to remember different images for each different website in order to achieve higher levels of security. If users cannot remember enough images and adopt only few images as passwords, the problem is similar to that of sharing textual passwords as discussed earlier. For biometric passwords, adversaries may get a user's biometric features from compromised websites; thus, they may be able to compromise a user's accounts on other websites as well. Furthermore, some biometric features, such as a user's fingerprints, are easy to obtain in the course of daily life.

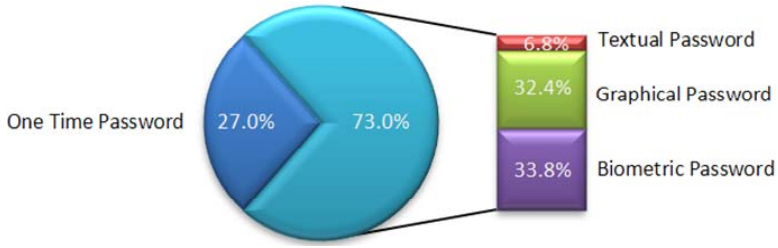
The one-time password scheme (RAS Laboratories, 2010), which is a dynamic password scheme and can achieve randomisation, is becoming popular and has been proved to be more secure. However, users need a device to receive an active password, such as smart cards (Sookyoung and Krishna, 2009) and mobile phones (Liao et al., 2009). Users have to carry these devices and use them to receive passwords when they are logging into the websites. It may not be convenient for some users. Furthermore, the security issues of such devices, such as being lost, need to be considered.

5.1.2 Users' trust for password schemes

In the survey, we also asked participants to choose the best password scheme based on convenience and safety of different password schemes. As shown in Figure 10, 27.0% of the participants believe one time password is best. 32.4% of the participants praise

graphical password scheme highly and 33.8% of the participants regard biometric password as the best. Only 6.8% of the participants still like textual password scheme. This result reflects that many participants (about 70%) prefer to use secure long time password schemes, such as graphic and biometric password schemes, which are easier to remember.

Figure 10 Password recovery schemes (see online version for colours)



5.2 Password management tools and single sign-on systems

5.2.1 Password management tools

Password management tools, such as Firefox extensions proposed by Halderman et al. (2005) to manage passwords, are considered to be potential solutions to the problem of using the same identities and passwords for multiple websites. Generally, a password management tool helps users to organise usernames and the corresponding passwords. Such a tool typically has a local database or file that holds the encrypted passwords for different websites. Users have to create profiles and set master passwords to manage their accounts' passwords. When users log into different websites, these tools can invoke corresponding passwords and input them into the login pages automatically. As such, users do not need to remember many passwords so that they can set unique and complex passwords for every website. We believe that these tools are convenient but they pose a high risk for all of users' accounts. This is because all accounts recorded in these tools will be compromised if the master passwords for these tools are compromised.

5.2.2 Single sign-on systems

Single sign on (SSO) systems allow users to log into their different accounts through one portal where the users are authenticated. Currently, there are two kinds of SSO schemes:

1 one password for different applications, such as Google applications and Microsoft online applications

2 one master password to manage SSO systems.

These SSO systems record users' identities and passwords in their databases. In SSO systems, users only need to use one account and a password to login; then, they can access all the websites listed without inputting identities and passwords again. For example, users just need to log into their Gmail account, and then they can access other services in Google such as Calendar, Doc, and Voice. SSO systems indeed improve the

level of convenience for users, but the security of all of the users' accounts depends on the security of the SSO systems – hence, they become single point of failure. Once a SSO system is compromised, all the accounts in it will be compromised.

5.2.3 Users' trust for password management tools and single sign-on systems

The results of our survey for password management tools demonstrate that 67.6% of the participants do not know anything about password management tools. Within the remaining 32.4% of the survey participants, 41.7%used these tools before and 62.5% believe that these tools are more convenient to manage passwords. However, 77.0% of the participants think these tools are not secure. The main reason is that these tools do bring the convenience to users but also they introduce higher risks to the users' accounts. Once the master password is compromised, all the accounts recorded in these tools will be compromised.

We also investigated participants' opinions about SSO systems. Our results show that all the participants know the SSO systems and 91.9% of the participants have used SSO systems before. 85.1% of the participants believe that SSO systems are more convenient. However, more than half of the participants cannot trust SSO systems. The reason why participants worry about SSO systems is the same as for the password management tools: it could be a single point of failure.

5.3 Recommendations for users

In previous sections, we discussed that users' behaviours, such as using the same e-mail address identities and passwords on multiple websites, put users' privacy at a higher risk and poses increasing security and privacy problems. There are some suggestions in the literature to overcome these related problems (Florencio and Herley, 2007; Florêncio et al., 2007), such as using diverse identities and applying password diversity. In the following, we provide some additional recommendations for users based on our findings:

- Avoid registering for an account using e-mail addresses on websites which have not been validated as secure and avoid using the same e-mail address as identity on many websites.
- Avoid using the passwords of e-mail accounts for accounts of other websites where these e-mail addresses have also been used as identities. Users should try to find their own password pattern, which is easy for them to remember, to achieve password diversity for multiple websites.
- A user should try to make his e-mail addresses and e-mail address identities independent of (or less dependent on) each other. For example, one improvement is to remove the e-mail account setting that allows forwarding from multiple accounts to just one. This mechanism can reduce the number of accounts compromised when a subset of the user's accounts are compromised.
- A user should go to the login pages with enhanced SSL technology if it is provided. For the websites that do not adopt a secure transmission schemes, it is not suggested to register accounts and transfer the sensitive information.

5.4 Future identity management systems

According to our analyses and discussions in the previous sections, a key issue of handling the problems related to e-mail address as identity is to protect users' private information on the websites, such as their passwords and answer to secret questions. A feasible solution is that users do not store their e-mail addresses, passwords and answers of secret questions on the websites. They should be able to provide their anonymous credentials when they log into various websites. A potential system employing this mechanism is the Federated Identity system where a trusted and secure third party is used to store users' personal information (Bertino et al., 2009). The cryptographic methods, such as zero knowledge protocols (Camenisch and Herreweghen, 2002), are used to guarantee anonymity and hence protect users' privacy. However, the high risk of exposing all of the accounts still exists. When a Federated Identity system is compromised by an adversary, all of the users' accounts on different websites are essentially compromised. Therefore, such a federated identity system must be well-designed to achieve a higher level of security.

Another promising identity management system is introduced by Schechter et al. in 2009. They employ a mechanism called social-authentication for identifying users. This system asks trustees (previously appointed by the account holder) to verify the account holder's identity. The main problem here is how to find appropriate trustees in a timely fashion when users want to log into the service. We believe that it may not be a significant problem in future due to the rapid development of smart phones and other internet-connected devices. In this case, the selected trustees may always stay connected with internet.

6 Conclusions and future work

Using e-mail address as identity on different websites is becoming a common practice. We believe that this trend is due to the convenience it provides to the users and websites that can use e-mail addresses to contact users for services. However, the use of an e-mail address as identity poses more security and privacy threats than username-based identity. We have demonstrated that the risks are due to several factors: use of e-mail addresses for multiple purposes, users' behaviours in using their e-mail address identities and passwords on multiple websites, and the design schemes used by websites which might not be secure. We have also analysed existing password management solutions and discussed two potential identity management systems. For future work, we plan to focus on improving and extending these two potential solutions to achieve more practical and secure identity management systems for dynamic large-scale web applications.

Acknowledgements

This research has been supported by the US National Science Foundation award IIS-0545912.

References

Amazon (2010) *Amazon Mechanical Turk*, available at https://www.mturk.com/mturk/ (accessed on 8/11/2010).

Bertino, E., Paci, F., Ferrini, R. and Shang, N. (2009) 'Privacy-preserving digital identity management for cloud computing'. *IEEE Data Eng*, Vol. 32, No.1, pp.21–27.

Camenisch, J. and Herreweghen, E.V. (2002) 'Design and implementation of the idemix anonymous credential system', in *Proceedings of the 9th ACM conference on Computer and Communications Security*, pp.21–30, 18–22 November, Washington, DC, USA.

DeadMan's Handle Ltd. (2005) *DeadMan's Handle and Cryptography*, available at http://www.deadmanshandle.com/papers/DMHAndCryptology.htm (accessed on 1/9/2010).

Dierks, T. and Allen, C. (1999) 'The TLS Protocol Version 1.0', RFC Editor, USA.

EffeTech (2010) *Ace Password Sniffer*, available at http://www.effetech.com/aps/ (accessed on 11/30/2010).

Everitt, K.M., Bragin, T., Fogarty, J. and Kohno, T. (2009) 'A comprehensive study of frequency, interference, and training of multiple graphical passwords', in *Proceedings of the 27th international conference on Human factors in computing systems*, pp.889–898, 4–9 April, Boston, MA, USA.

Florencio, D. and Herley, C. (2007) 'A large-scale study of web password habits', Paper presented at the *16th international conference on World Wide Web*, pp.657–666, 8–12 May, Banff, Alberta, Canada.

Florêncio, D., Herley, C. and Coskun, B. (2007) 'Do strong web passwords accomplish anything?', in *Proceedings of the 2nd USENIX Workshop on Hot Topics in Security*, pp.1–6, 7 August, Boston, MA, USA.

Halderman, J.A., Waters, B. and Felten, E.W. (2005) 'A convenient method for securely managing passwords', in *Proceedings of the 14th International Conference on World Wide Web*, pp.471–479, 10–14 May, Chiba, Japan.

James, L.W. (2008) 'Biometrics in identity management systems', *IEEE Security and Privacy*, Vol. 6, No.2, pp.30–37.

Jin, L. (2010) *Investigations of Users' and Websites' Behaviors*, available at www.sis.pitt.edu/~leijin/investigation.htm (accessed on 1/9/2010).

Jin, L., Takabi, H. and Joshi, J.B.D. (2010) 'Security and privacy risks of using e-mail address as an identity', in *Proceedings of the Second International Conference on Social Computing*, August 20–22, Minneapolis, Minnesota, USA.

Just, M. and Aspinall, D. (2009) 'Personal choice and challenge questions: a security and usability assessment', in *Proceedings of the 5th Symposium on Usable Privacy and Security*, pp.1–11, 15–17 July, Mountain View, CA, USA.

Liao, K.C., Lee, W.H., Sung, M.H. and Lin, T.C. (2009) 'A one-time password scheme with QR-code based on mobile phone', in *Proceedings of the 5th International Joint Conference on INC, IMS and IDC*, pp.2069–2071, 25–27 August, Seoul, Korea.

Perlman, R. and Kaufman, C. (2008) 'User-centric PKI', in *Proceedings of the 7th Symposium on Identity and Trust on the Internet*, pp.59–71, 4–6 March, Gaithersburg, MD, USA.

RAS Laboratories (2010) *One-Time Password Specifications*, available at http://www.rsa.com/rsalabs/node.asp?id=2816 (accessed on 11/29/2010).

Schechter, S., Egelman, S. and Reeder, R.W. (2009) 'It's not what you know, but who you know: a social approach to last-resort authentication', in *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, pp.1983–1992, 04–09 April, Boston, MA, USA.

Sookyoung, L. and Krishna, M.S. (2009) 'An efficient one-time password authentication scheme using a smart card', *Int. J. Secur. Netw*., Vol. 4, No. 3, pp.145–152.

Zheleva, E. and Lise, G. (2009) 'To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles', in *Proceedings of the 18th International Conference on World Wide Web*, pp.531–540, April 20–24, Madrid, Spain.