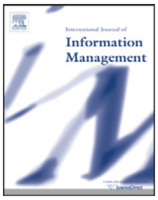
Contents lists available at SciVerse ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt



Improving information security management: An analysis of ID–password usage and a new login vulnerability measure

Youngsok Bang[a], Dong-Joo Lee[b], Yoon-Soo Bae[c], Jae-Hyeon Ahn[c,*]

[a] Desautels Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, Quebec, Canada
[b] Division of Management, Hansung University, 102 Hansungdae Street, Sungbook-gu, Seoul, Republic of Korea
[c] KAIST Business School, 207-43 Chongyangri-dong, Dongdaemoon-gu, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Available online 18 February 2012

Keywords:
Information security management
ID
Password
Vulnerability measurement
E-business

ABSTRACT

Statistics show that the number of identity theft victims in the US increased by 12% in 2009, to 11.1 million adults, while the total annual fraud amount increased by 12.5%, to $54 billion. As the e-commerce volume is increasing and various online services are becoming more popular, the number of sites to which an average Internet user subscribes is increasing rapidly. Given the limited memory capacity of human beings, an Internet user's login credentials (in the form of a combination of a user ID and a password) are usually reused over multiple accounts, which can cause significant security problems. In this study, we address the vulnerability of login credentials. First, based on a unique Internet user data set, we analyze the behavioral characteristics of login credentials usage. We find that the same login credentials are used for many more accounts and reused much more often than previously expected. Furthermore, usage patterns are found to be quite skewed. Second, building on a network perspective of login credentials usage, we suggest a vulnerability measure of an individual's login credentials and analyze the vulnerability of current Internet users. The resulting information is valuable not only to the research community but also to managers and policy makers striving to reduce security vulnerability.

1. Introduction

A guiding tenet of information security is that security is only as strong as the weakest link and users are the weakest link (Schneier, 2000). As such, information security is not only a technical issue but also a behavioral issue involving users. Much research has been conducted to understand users' security-related behaviors, such as information systems misuse (D'Arcy, Hovav, & Galletta, 2009; Siponen & Vance, 2010; Straub, 1990; Workman, Bommer, & Straub, 2008) or security-enhancing actions (Bulgurcu, Cavusoglu, & Benbasat, 2010; Johnston & Warkentin, 2010; Kankanhalli, Teo, Tan, & Wei, 2003; LaRose, Rifon, & Enbody, 2008) mostly in work environment settings.

Unlike employees in a work environment setting, however, general end users are not subject to training, nor are they protected by a technical security staff at work. Thus, with over a billion people with access to the Internet, individual Internet users represent a significant point of weakness in cybersecurity (Anderson & Agarwal, 2010). As e-commerce volume continues to expand and various online services—including e-mail, financial, social networking, and content services—become increasingly popular, the number of sites to which an average Internet user subscribes is increasing rapidly, generating a significant security issue over multiple accounts.

To gain access to a website account, each user usually has to go through an identification and authentication process (Pernul, 1995). The most prevalent website identification/authentication mechanism is the use of credentials in the form of a user ID/password (PW) combination (hereafter referred to as login credentials). Security-enhancing measures generally fall into four distinct sequential activities called the Security Action Cycle—deterrence, prevention, detection, and recovery (Straub & Welke, 1998). The identification/authentication mechanism is representative of prevention activities in the cycle (Doherty, Anastasakis, & Fulford, 2011; Kankanhalli et al., 2003; Straub & Welke, 1998).

Since multiple accounts exist, the reuse of login credentials (the same combination of an ID and a PW) for accessing more than one account can cause serious problems, as has been widely suggested in the literature (e.g., Adams & Sasse, 1999; Gaw & Felten, 2006; Ives, Walsh, & Schneider, 2004; Zhang, Luo, Akkaladevi, & Ziegelmayer, 2009). For example, a security breach on one site can trigger a security risk on other sites, because a hacker who gains access to one account may be able to gain access to others (Ives et al., 2004). In fact, users who reuse PWs often fail to realize that

[*] Corresponding author. Tel.: +82 2 958 3677; fax: +82 2 958 3667.
E-mail addresses: youngsokbang@gmail.com (Y. Bang), dongjoo2@gmail.com (D.-J. Lee), bluebys@business.kaist.ac.kr (Y.-S. Bae), jahn@business.kaist.ac.kr (J.-H. Ahn).

0268-4012/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijinfomgt.2012.01.001

their most well-defended account is not more secure than their most poorly defended account due to the reuse (Ives et al., 2004). The effect of the resulting identity fraud can be substantial. Statistics show that the number of identity fraud victims in the US in 2009 increased by 12%, to 11.1 million adults, while the total annual fraud amount increased by 12.5%, to $54 billion (Javelin Strategy and Research, 2010).

A recent case in South Korea provides a clear example of the problem of a security breach and the subsequent crime committed by using reused login credentials. In 2008, two hackers attacked about 100 small, less secure websites in South Korea—such as flower delivery sites, online game sites, and real estate sites—and stole the login credentials of 2.3 million users. Using these login credentials, they hacked into Naver.com, which is the most popular portal site in South Korea commanding over 60% of the market share. As many as 150,000 accounts in the portal site were successfully attacked. The hackers exploited the account information for fraudulent advertising and sold the personal information acquired from the accounts. While Naver.com is generally considered to have strong system security, this case shows that the company's user data are no longer secure because of the company's links with other less secure sites in which the login credentials of users are the same as those in Naver.com. Therefore, it is reasonable to expect that the risks caused by reusing login credentials will increase exponentially, because the number of systems protected by login credentials (particularly small websites) is increasing (Ives et al., 2004).[1]

Cognitive psychology theory provides an explanation for the reuse behavior. It argues that human beings have an inherently limited memory capacity (Miller, 1994). Given multiple accounts, users must perform a mental process of searching and retrieving the account-login credential pairs from their memory (Zhang et al., 2009). Because of the memory problem, remembering and managing multiple IDs and PWs becomes difficult and cumbersome.

The starting point for addressing the vulnerability of login credentials is to understand the status of reuse behavior. However, it is difficult to obtain even a single firm's data about users' actual login credentials. Therefore, objective and comprehensive statistics are rarely available at present. In fact, most statistics about reused credentials from previous studies (e.g., Gaw & Felten, 2006; Kaspersky Lab, 2007; RSA, 2004) are based on the respondents' *speculation* about their accounts and their usage of login credentials in the accounts rather than on *objective* data. However, recall may not be reliable. When users subscribe to many sites, they may fail to recall some of the login credentials used and even forget some of the sites in general. Our results show that current recall-based reuse statistics are usually quite biased.

The study by Florencio and Herley (2007) is exceptional because it is based on large-scale, objective data on PW reuse, gathered over 10 weeks through a component of Windows Live Toolbar. However, as the authors indicate, the study may have missed a large fraction of PW usage behavior, because users may log into their accounts from more than one machine, which could not be accounted for in the data-gathering method, and PWs with a bit strength of less than 20 were not included in the collection. Furthermore, PWs used at only one site were excluded, because data were collected only for reused PWs. In addition, users may not have logged into some infrequently visited sites during the 10-week observation period.

Therefore, our knowledge of the current state of the reuse of login credentials appears to be limited. To overcome this problem, we provide and analyze objective and comprehensive statistics based on a unique data set about the reuse of login credentials. This is the first objective of this study.

The second objective of the study is to suggest a measure of the vulnerability of login credentials and analyze the data to assess the vulnerability of current Internet users. An appropriate measure of vulnerability can be used to assess related risks and can guide the allocation of resources for security improvement (Alhazmi, Malaiya, & Ray, 2007). Although several measures of login credentials vulnerability—such as PW strength, memorability, and the PW reuse ratio—have been applied, they either ignore login credentials reuse over multiple accounts (e.g., PW strength and memorability) or do not reflect the structural characteristics of the reuse (e.g., the PW reuse ratio).

PW strength, which measures the effectiveness of a single PW in preventing guessing and brute-force attacks, presumes that the account is well defended if the PW is long, complex, and unpredictable (Burr, Dodson, & Polk, 2006; Horcher & Tejay, 2009; Weber, Guster, & Safonov, 2008). Therefore, a random sequence of upper- and lower-case letters, punctuations, symbols, and numbers are typically used to generate an ideal PW. However, as we can see from the identity theft case of Naver.com, the vulnerability of an account depends not only on the security level of the account itself, but also on the behavioral patterns of login credentials reuse. In addition, a strong PW tends to be difficult to remember and that may lead to a security problem because a user might keep an insecure written record of it or rely on an insecure backup authentication procedure after forgetting it (Yan, Blackwell, Anderson, & Grant, 2004). Given this PW strength–memorability tradeoff, PW memorability, which measures the ease with which the user can remember a PW, has drawn research attention to examine efficient ways to improve memorability without compromising strength (Bunnell, Podd, Henderson, Napier, & Kennedy-Moffat, 1997; Nelson & Vu, 2010; Vu et al., 2007; Yan et al., 2004). However, similar to PW strength, PW memorability has a focus on a single PW, without considering login credentials usage over multiple accounts. Finally, the PW reuse ratio, another popular measure of login credentials vulnerability (e.g., Brown, Bracken, Zoccoli, & Douglas, 2004; Florencio & Herley, 2007; Gaw & Felten, 2006), is defined as the number of sites to which a user subscribes divided by the number of unique PWs used at the sites. As long as a given number of unique login credentials are used over a given number of accounts, the reuse ratio is identical, regardless of the structural characteristics of the reuse, that is, how the login credentials are distributed over the accounts. Therefore, the reuse ratio is most appropriate when the login credentials are uniformly distributed, while it is subject to a bias when applied to a skewed usage pattern of login credentials, which is shown to be the case in this paper.

These limitations of the current measures suggest the need for a new measure of vulnerability that captures the behavioral patterns and structural characteristics of the login credentials usage over multiple accounts.[2] To fill this gap, we propose a measure of the login credentials vulnerability that can be applied in practice and have a clearly defined interpretation.

The rest of the study is organized as follows. Section 2 details the method of data collection. Section 3 presents the major findings from the analysis of login credentials usage and compares the statistics from our data with those from previous studies. Section 4 suggests a network perspective on the usage of login credentials to clearly understand the characteristics of usage patterns of login

[1] In April 2011, one of the largest recorded data breaches occurred at Sony's computer networks. Over 100 million accounts were compromised from PlayStation Network, Qriocity, and Sony Online Entertainment Network. Stolen account information included IDs, PWs, names, addresses, etc. (McMillan, 2011).

[2] Other literature has suggested methods to measure the vulnerability of IT systems (e.g., Farahmand, Navathe, Sharp, & Enslow, 2005; Patel, Graham, & Ralston, 2008; Wang, Wang, & Wulf, 1997). However, system-level measures are hard to apply for the assessment of the vulnerability of login credentials at the individual user level.

credentials and the resulting vulnerability. Section 5 proposes a new vulnerability measure reflecting the characteristics of login credentials usage. Finally, in Section 6, some discussions and concluding remarks are presented.

2. Data collection

Our main sample consists of 49 Internet users in South Korea.[3] To collect their login credentials for the different sites that they have subscribed to as thoroughly as possible, we used an Internet security site (www.sitechecker.co.kr), which searched about 30,000 South Korean sites and provided a list of all the sites to which an Internet user had subscribed on the basis of the user's identification information (name and social security number, or SSN). This is possible because users' real names and SSNs are required and verified by almost all South Korean websites upon sign-up. This unique feature makes South Korea an excellent region for obtaining reliable, although not perfect, data for the study.

The participants were asked to enter their identification information into the site to obtain a list of the sites to which they were subscribing. Then, they were asked to provide their IDs and PWs for the sites using a serial number (e.g., ID1, ID2,. . .; PW1, PW2,. . .); the same serial number for IDs in two sites indicated the same ID for the sites, and similarly for PWs. To ensure the accuracy of the data, every data collection session was conducted face-to-face and each respondent was asked to log into all the sites retrieved. Upon failure to recall either an ID or a PW for a site, the respondent was asked to inquire the site for correct information on ID or PW. Thus, we could verify whether each respondent's IDs (PWs) for any pair of sites were the same. We offered the participants some financial rewards. It usually took about 1–2 h to complete the data collection for each participant. Since the data collected were very private, it was difficult to get a large sample size. There were 34 (69%) male and 15 (31%) female respondents; 32 (65%) respondents were in their 20's, 11 (23%) were in their 30's, three (6%) were in their 40's, and the remaining three (6%) were in their 50's. Of all the respondents, 18 (37%) were undergraduate or graduate students, and 31 (57%) had full-time jobs while three (6%) were housewives.

3. Analysis of login credentials usage

This section analyzes the data to understand login credential usage behaviors over multiple accounts and contrasts the results with those from previous recall-based studies. Table 1 summarizes the main descriptive statistics from our data analysis. We provide the major findings below.

3.1. Finding 1: the number of subscribing accounts is considerably larger than previously expected.

The statistics on the number of subscribing accounts obtained using our unique data set are substantially different from those obtained in previous studies. This difference causes a great difference in the reuse ratio estimates, as explained subsequently. The average number of accounts is 105.7 (median = 95), ranging from 27 to 199 (see Table 1, first row). The average number is considerably larger than those in the existing statistics summarized in Table 2, where the average or median values are mostly less than 10 and at most 25. For example, Gaw and Felten (2006) found that the average number of accounts was only 7.9. In their study, the 49 participants were asked to indicate the websites they used out of 139 sites. Further, they were requested to recall and add other sites at which they had their own accounts. Another recall-based survey of 150 users in the UK by Kaspersky Lab (2007) revealed that 62% of users have 10 or fewer online accounts with PWs and that only 23% of users have more than 20 accounts. Brown et al. (2004) surveyed 218 college students and reported a similar result: Each student had on, an average, 8.2 PW-protected accounts.

The substantial discrepancy between our results and those of existing studies seems to be closely related with the difference in data-gathering methods (i.e., objective versus recall based).[4] Alternatively, the discrepancy may have been caused by the difference in Internet usage among different users of different countries. However, related statistics do not show any significant difference in Internet usage among the users in South Korea, the US, and the UK: As of 2010, Internet penetration rates, the number of Internet users out of total population, were 81.1%, 77.3%, and 82.0%, respectively, in the three countries (Miniwatts Marketing Group, 2011). The usage rates of the users for major Internet services are also similar between South Korea and the US: 87.8% and 91.0% for services, 61.3% and 71.0% for Internet shopping, 45.0% and 55.0% for Internet banking, and 86.7% and 70.0% for Internet news services, respectively (National Internet Development Agency of Korea, 2009). As of 2009, the average Internet user from South Korea and the US spent an estimated 2.0 h and 2.1 h online per day, respectively, and 2.1 h and 1.9 h as of 2010 (Korea Communications Commission and Korea Internet & Security Agency, 2009, 2010; The Nielsen Company, 2010a, 2010b). Thus, the difference in Internet usage among countries is not likely to be the source of the discrepancy in the analysis results.

Given the small number of respondents in our main sample, we gathered another set of data (hereafter referred to as the *supplementary data set*) to demonstrate that recall-based surveys of login credentials usage tend to significantly under-report the number of accounts used. The data were collected from 50 undergraduate students enrolled in an information systems class at a South Korean university who did not overlap with the respondents in the main sample.

Both the recall-based method as in the previous studies and the objective method as in Section 2 were applied to each respondent. Specifically, each respondent was first asked to recall the websites to which he or she was subscribing and provide the estimate of the total number of the sites. Next, each respondent was asked to retrieve a list of the subscribing sites from the Internet security site www.sitechecker.co.kr and provide the total number of retrieved sites. This approach enabled us to analyze the effect of the data-gathering method while controlling for other confounding factors such as Internet usage or environmental differences. We also gathered data about the respondents' reactions to periodic PW change requests from websites, which will be discussed later.

The analysis of the supplementary data shows that the results are consistent with those from the main sample. First, the average number of sites retrieved from the Internet security site is 122.5 (median = 119.5). This number is not statistically different from the average number of sites in the main sample (105.7), with $t = 1.523$ ($p = 0.131$), implying that the statistics in Table 1 are likely to be robust across both samples. Second, the average number of sites recalled by the respondents is 50.0 (median = 31.0). A pairwise t-test verifies that the number of retrieved sites is significantly larger than the number of recalled sites, with $t = 9.048$ ($p < 0.001$).

The above results from our two samples suggest that Internet users have difficulty recalling the sites to which they subscribe.

[3] We also gathered data from another sample, which is described in Section 3.

[4] Our result is also quite different from that of Florencio and Herley (2007), who found on the basis of large-scale, objective data that the average number of accounts of an Internet user is 25. This seems to be related to the potential sources of error in the study, as mentioned before.

Table 1
Status of login credentials usage.

Item	Mean	S.D.	Min	Median	Max
No. of sites that users subscribe to (accounts)	105.7	42.9	27	95	199
No. of unique IDs used	6.6	3.1	2	6	14
No. of unique PWs used	4.7	2.3	1	4	15
No. of unique (ID, PW) combinations used	11.8	5.3	4	11	28
ID reuse ratio[a]	19.1	11.7	4.3	16	68.5
PW reuse ratio[a]	29.2	29.2	3.9	23	199
(ID, PW) reuse ratio[a]	10.5	6.9	2.1	8.5	39.8
Percentage of active (ID, PW)s[b]	45.6%	16.7%	13.3%	42.9%	100.0%

[a] The ID (PW) reuse ratio is the number of sites that users subscribe to divided by the number of unique IDs (PWs) used; the (ID, PW) reuse ratio is the number of sites that users subscribe to divided by the number of unique (ID, PW) combinations used.

[b] The percentage of active (ID, PW) combinations is the number of unique (ID, PW) combinations used divided by the product of the numbers of unique IDs and PWs used).

They usually substantially underestimate the number and therefore may underrate the potential risks in reusing login credentials. Hence, recall-based studies have significant limitations in providing a credible picture of login credentials usage.

3.2. Finding 2: the same IDs and PWs are very frequently used for multiple accounts

Table 1 shows that the respondents use only a small number of unique PWs, that is, 1–15 PWs, with an average of 4.7 (median = 4) PWs. This result is similar to those of previous studies (see Table 2), where the average numbers range from 3.3 to 6.5 and the medians are less than 5.

Using the data on the number of the sites to which each respondent subscribes, we can compute the PW reuse ratio for respondents, which is defined as the number of sites to which a user subscribes divided by the number of unique PWs used at the sites. As shown in Table 1, the average of the ratios is 29.2 (median = 23). This contrasts sharply with the average reuse ratios reported in previous studies, which range from 1.8 to 3.9 (see Table 2). Therefore, since the Internet users' PWs are reused intensively over multiple accounts, the resulting vulnerability is considerably greater than what was previously expected.

Our study also provides statistics on ID usage, which have been rarely reported to date. The respondents use 6.6 unique IDs on average (median = 6), with the range being 2–14. The average ID reuse ratio is 19.1, which is also very high. By relating the ID usage data to the PW usage data, we find that the respondents use a higher number of unique IDs than unique PWs ($t = 3.997$, $p < 0.001$). In addition, the correlation coefficient (0.106) between the number of unique IDs and the number of unique PWs is not significant ($p = 0.472$) when one unusual observation with 14 unique IDs and 15 unique PWs is excluded. That is, users maintaining more diversity in their IDs do not necessarily use more PWs; rather, both are independent. Since Internet users employ multiple IDs as well as multiple PWs and since both IDs and PWs are highly reused over a larger number of accounts, the management of login credentials and the corresponding risks need to be examined from the perspective of a combination of the two; this perspective is adopted in the following analysis.

3.3. Finding 3: only a limited proportion of the possible (ID, PW) combinations is actually used. Furthermore, the reuse ratio of (ID, PW) combinations is higher for users with more accounts

By replacing IDs or PWs in the above analysis with combinations of IDs and PWs—denoted (ID, PW)—we find that the respondents use 11.8 unique (ID, PW) combinations, on average. Therefore, they use combinations in a more diversified manner compared to just IDs or PWs (see the second to fourth rows in Table 1). However, the reuse ratios of the combinations are still very high, with an average of 10.5 (the seventh row in Table 1).

An interesting finding is that the respondents do not diversely use the possible combinations of IDs and PWs. Consider a user who has six IDs and five PWs; 30 unique combinations of IDs and PWs are then possible. However, the respondents use only 45.6% of the possible combinations, on average (see Table 1, last row). Thus, the reuse ratio of the active (ID, PW) combinations increases. This result implies that Internet users may substantially mitigate the problems of reusing login credentials by diversifying the combinations of their current IDs and PWs without using additional IDs or PWs.

Another related observation is that the number of accounts and the number of combinations are not significantly correlated, with a correlation coefficient of 0.164. Further, the reuse ratio of combinations is found to have a strong positive correlation with the number of accounts (correlation coefficient = 0.648, $p < 0.001$). Therefore, subscription to more sites is not usually accompanied by diverse (ID, PW) combinations, leading to a higher reuse ratio. This result can also be attributed to the cognitive limitations of users in managing their login credentials over multiple sites.

Table 2
Statistics from previous studies.

Study	Account type	Data-gathering method	No. of unique accounts	No. of unique PWs	Reuse ratio
Gaw and Felten (2006)	Websites	A recall-based survey of 49 respondents	Mean = 7.9	Mean = 3.3	3.2
Kaspersky Lab (2007)	Websites	A recall-based survey of 150 users in the UK	≤10: 62%≤20: 15%≥21: 23%	≤4: 51%≤10: 30%≥11: 19%	Not reported
Florencio and Herley (2007)	Websites	Observation of half a million users	Mean = 25	Mean = 6.5	3.9
Brown et al. (2004)	Websites, credit card, computer systems, etc.	A recall-based survey of 218 students in the US	Mean = 8.2	Mean = 4.5	1.8
RSA (2004)	Websites, computer systems, ATMs, etc.	A recall-based survey of 1022 adults in the US	Not reported	≤4: 63% ≥5: 37%	Not reported

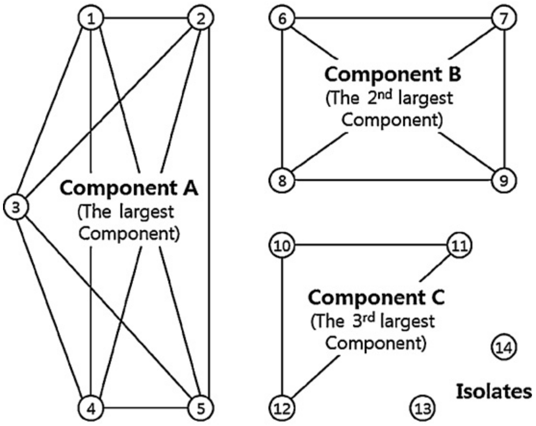


Fig. 1. An illustration of a user's ID–PW usage network (with three components and two isolates).

4. A network perspective to login credentials vulnerability

The previous section, using the measure of reuse ratio, contrasted login credential usage statistics from our study with those from recall-based studies and found that recall-based data tend to generate substantially biased results on login credential usage behaviors. This section carries the analysis a step further to examine the usage structure of login credentials and thereby show the limitation of the reuse ratio in capturing security vulnerability.

Our analysis is based on network theory. Euler (1741) laid the foundation for network theory by introducing graph concept. A graph consists of points (called nodes), a set of discrete elements, and lines (called links), a set of connections between pairs of points. These points and lines concepts could be almost anything: people and friendships (Rapoport & Horvath, 1961), computers and communication lines (Faloutsos, Faloutsos, & Faloutsos, 1999), chemicals and reactions (Jeong, Tombor, Albert, Oltvai, & Barabasi, 2000; Wagner & Fell, 2001), scientific papers and citations (Redner, 1998), and journal authors and their joint papers (Goldenberg, Libai, Muller, & Stremersch, 2010). The network perspective abstracts away all the details of the real problem, focusing on the structure of connectivity (relationship). Recently, network theory has been widely applied to business research such as user–IT system interactions (Kane & Alavi, 2008), advertising competition (Chang, Oh, Pinsonneault, & Kwon, 2010), knowledge diffusion (Blumenberg, Wagner, & Beimborn, 2009; Hansen, 2002; Janhonen & Johanson, 2011), and new service development (Syson & Perks, 2004).

By applying network theory to login credentials usage, we can easily capture how an Internet user manages her login credentials for her subscribing accounts. Specifically, the Internet sites to which a user subscribes can be modeled as a network in which each site is viewed as a node. A link between two sites is created if the user uses the same login credentials on both sites. Thus, the link between the sites maps the transmission of vulnerability caused by the reuse of login credentials.

Fig. 1 illustrates a network representation of a hypothetical user (ID, PW) usage with five (ID, PW) combinations over 14 sites. The combinations are used, respectively, on five sites (Sites 1–5), four sites (Sites 6–9), three sites (Sites 10–12), one site (Site 13), and one site (Site 14). Each site is represented by a node with the corresponding number. Because the user uses the same combination on Sites 1–5, Nodes 1–5 are linked together: these nodes constitute a *component* (denoted by Component A in Fig. 1). In network theory, a component is defined as a maximal connected subnetwork, that is, a subnetwork of the nodes that are linked between themselves but not linked outside to other nodes (Nooy, Mrvar, & Batagelj, 2005). Similarly, Nodes 6–9 on which another combination is shared constitutes another component (Component B). Component C is constructed the same way, with three nodes (Nodes 10–12). If an (ID, PW) combination is used only for one site, the corresponding node has no link and becomes an *isolate* (in network theory terminology). In Fig. 1, both Nodes 13 and 14 are isolates. Thus, the number of (ID, PW) combinations used by a user is equal to the sum of the number of isolates and the number of components in the corresponding network.

Using this approach, we can derive, for each respondent, an *ID–PW usage network* of the sites to which the respondent subscribes and apply network theory to investigate its structural characteristics. Table 3 summarizes the results.

4.1. Finding 4: the usage patterns of login credentials are highly skewed

In network theory, the *inclusiveness* of a network is defined as the number of connected nodes expressed as a proportion of the total number of nodes (Nooy et al., 2005). The respondents' ID–PW usage networks have an average inclusiveness of 0.94 (see Table 3, first row). This means that for an average user, 94% of the sites to which the user subscribes have the same login credentials as at least one other site, which results in potential security breach chains, and only 6% of sites are isolated in terms of the security risk of login credentials.

Given the high level of inclusiveness, a question follows concerning the distribution of (ID, PW) combinations over sites with connections, that is, the distribution of the size of the components. A common measure of the ratio of the *k*th largest component to the entire network, or the number of sites in the *k*th largest component to the total number of sites in the network (Goldenberg et al., 2010), provides relevant information, as shown in Table 3.

The ratio of the largest component to the entire network is 0.54 (average); that is, the most frequently used combination for each respondent is used for almost 54% of the total sites to which the respondent subscribes. Therefore, if the login credentials are stolen, for example, while signing up at a fake site, more than half of the total accounts are potentially at risk. In an extreme case, one respondent had used a single combination for over 87% of her total accounts. The average ratios for the second and third largest components are 0.18 and 0.09, respectively. Therefore, the three most frequently used combinations of each respondent are used for an average of 81% of the respondent's accounts. By comparing this result with the average number of unique combinations, 11.8, in Table 1, we can see that Internet users' usage patterns of their login credentials are highly skewed. They use very few combinations for most sites.

The highly skewed nature of login credentials usage implies an inherent limitation of the reuse ratio as a measure of vulnerability. For a given number of accounts and a given number of unique (ID, PW) combinations, the reuse ratio is identical by definition, independent of whether the usage is uniform or skewed over the accounts. However, both usage patterns are not subject to the same level of vulnerability because the severity of a potential breach would be affected by the degree of skewness of the (ID, PW) usage network, as shown in the following section. Thus, we suggest a new measure of vulnerability that considers the structure of the network and captures the vulnerability caused by the skewness.

The above findings are based on our sample data. They may need to be interpreted with caution, since the sample is relatively small and not from random sampling, mainly due to the highly private nature of the data-gathering method and the significant effort required for response. However, the results can serve as an important starting point on why a new vulnerability measure is needed.

Table 3
Structural characteristics of ID–PW usage networks.

Item	Mean	Standard deviation	Min	Median	Max
Inclusiveness	0.94	0.05	0.74	0.95	1.00
Ratio of the largest component to the entire network	0.54	0.19	0.20	0.56	0.87
Ratio of the second largest component to the entire network	0.18	0.08	0.04	0.17	0.40
Ratio of the third largest component to the entire network	0.09	0.04	0.02	0.09	0.21
Vulnerability index (VI)	0.38	0.18	0.07	0.37	0.75

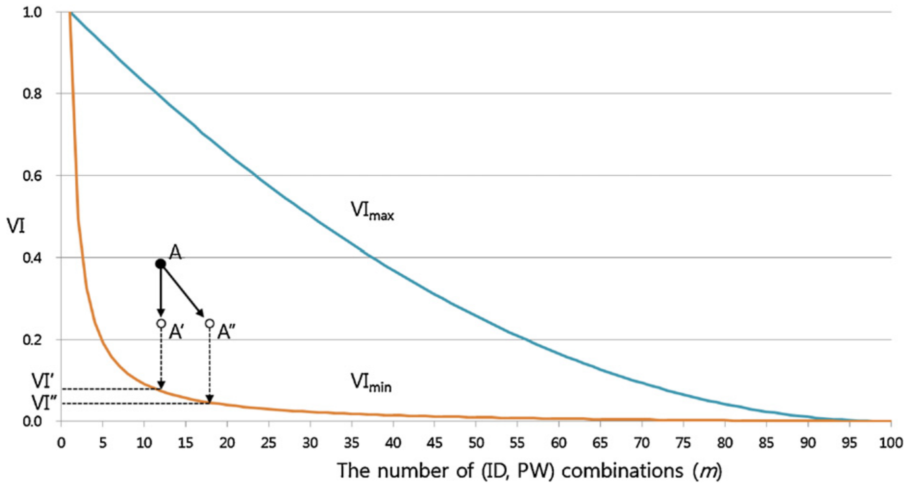


Fig. 2. Illustration of the VI.

5. Vulnerability of login credentials: a measure and analysis

In this section, we suggest a useful and informative measure of the vulnerability of login credentials, termed the *vulnerability index* (VI), and analyze the data. Suppose that a user is subscribing to N sites and uses m (ID, PW) combinations on the sites. Let c_i denote combination i ($i = 1, 2, \ldots, m$) and n_i denote the number of sites where combination i is used ($n_1 + n_2 + \cdots + n_m = N$). Consider an extreme case in which the user uses unique combinations for the sites (i.e., $m = N$). Then, all the nodes become isolates and no link exists. A security breach at any one site would not make the login credentials for other sites vulnerable. Thus, this is the most secure case.

In the other extreme case, the user may use only one combination across all sites ($m = 1$). Then, all the pairs of nodes are linked, generating ${}_NC_2$ links. A breach at any one site would make the login credentials for all remaining $N - 1$ sites vulnerable. Thus, this is the least secure case.

In an intermediate case ($1 < m < N$), at least one component should exist and isolates may exist. If a breach occurs in one site, the other sites in the same component are exposed to the risk. On the other hand, a breach at any isolated site does not harm the other sites. Thus, the severity of a potential breach depends on the structure of the network and the site of the initial breach.

Using this observation, the VI of an ID–PW usage network is defined as the expected proportion of sites subject to potential breaches if a breach at one site occurs. Suppose that $N = 6$, $m = 3$, $n_1 = 3$, $n_2 = 2$, and $n_3 = 1$. Assume that the probability of being a victim of the initial breach is the same for all sites. Then, given a breach at one site, the login credentials for the site would be c_1 with probability 3/6, c_2 with probability 2/6, or c_3 with probability 1/6. If the login credentials are c_1, a breach risk involving c_1 exists at two additional sites. If the login credentials are c_2, a breach risk involving c_2 exists at one additional site. If the login credentials are c_3, no additional breach risk exists. Thus, the expected proportion of vulnerable sites is equal to $(3/6) \times (2/5) + (2/6) \times (1/5) + (1/6) \times (0/5) = 0.27$, which implies that one successful breach could cause breaches at 27% (average) of the remaining sites. By a simple generalization, we obtain the following formal expression:

$$\mathrm{VI} = \sum_{i=1}^{m} \left(\frac{n_i}{N}\right)\left(\frac{n_i - 1}{N - 1}\right) \tag{1}$$

To derive Eq. (1), we assume the same probability of being the victim of the initial breach for all N sites. However, it can be shown analytically that the VI formula is valid even in the presence of differences in probability.

It is easy to verify that VI = 0 when $m = N$ and VI = 1 when $m = 1$. A larger value of VI indicates a higher level of vulnerability. For a given N and m, it can be shown that VI increases with the variance of n_i.[5] Thus, VI is minimum ($\mathrm{VI}_{\min}$) when the variance of n_i is zero; that is, all the combinations are used on the same number of sites (i.e., $n_1 = n_2 = \cdots = n_m = N/m$).[6] In this case, $\mathrm{VI}_{\min} = (N - m)/m(N - 1)$ (from footnote 4). The variance of n_i would be highest when $n_i = 1$ for $i \neq j$ and $n_j = N - (m - 1)$, and in this case VI will be maximum ($\mathrm{VI}_{\max}$) and $\mathrm{VI}_{\max} = (N - m + 1)(N - m)/N(N - 1)$ from Eq. (1).

The VI proposed does not distinguish the relative importance of ID and PW. ID is known publically in some cases because it may be

[5] $$\mathrm{VI} = \sum_{i}^{m} \left(\tfrac{n_i}{N}\right)\left(\tfrac{n_i-1}{N-1}\right) = \tfrac{m}{N(N-1)} \sum_{i}^{m} \left(\tfrac{n_i^2}{m} - \tfrac{n_i}{m}\right)$$
$$= \tfrac{m}{N(N-1)} \left[\sum_{i}^{m} \tfrac{n_i^2}{m} - \left(\sum_{i}^{m} \tfrac{n_i}{m}\right)^2 + \left(\sum_{i}^{m} \tfrac{n_i}{m}\right)^2 - \sum_{i}^{m} \tfrac{n_i}{m}\right] =$$
$$\tfrac{m}{N(N-1)} \left[\mathrm{Var}[n_i] + \left(\tfrac{N}{m}\right)^2 - \tfrac{N}{m}\right] = \tfrac{m}{N(N-1)} \mathrm{Var}[n_i] + \tfrac{N-m}{m(N-1)}$$

[6] Since every n_i is an integer, the minimum variance would be larger than zero if N/m is not an integer. Thus, $\mathrm{VI}_{\min}$ would be larger than $(N - m)/m(N - 1)$. However, the gap, $m\mathrm{Var}[n_i]/N(N - 1)$, is negligible, given that N is sufficiently large and $\mathrm{Var}[n_i]$ is close to zero.

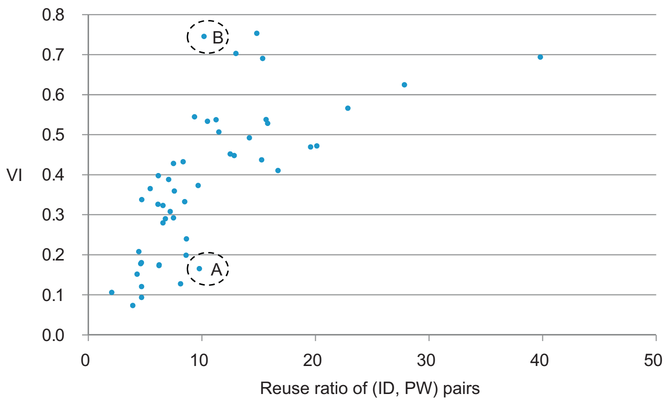


Fig. 3. Respondents' reuse ratio and VI values.

included in an or used as a nickname for a community. Considering the publicity of ID, as a supplement, we can calculate the VI on the basis of PW usage only, which can be easily derived analogously to the VI proposed here.

Fig. 2 illustrates the minimum and maximum VI values as a function of m for $N = 100$. As m increases, both VI_{min} and VI_{max} decrease. Suppose that point A represents the current VI, 0.4, of a user's ID–PW usage network. The user can decrease the VI (e.g., from A to A′) by reducing the variation in the number of sites where each combination is used (i.e., by decreasing the variance of n_i). Alternatively, the user can decrease the VI (e.g., from A to A″) by using more combinations (m). In either case, the user can further decrease the VI (to VI′ or VI″) by minimizing the variance of n_i. Thus, VI provides information about not only the current level of vulnerability but also the possible extent of reduction in vulnerability.

We calculate the VI values for the 49 respondents. As shown in Table 3 above, the average VI is quite high (0.38); therefore, if a breach occurs at any one of the sites to which a respondent subscribes, on average, 38% of the remaining sites could experience potential breaches. Fig. 3 links the reuse ratios and the VI values for the respondents. It clearly shows that substantial variations exist in the VI values of respondents with similar reuse ratios. For example, respondents A and B have similar reuse ratios of about 10, but their VI values (0.17 and 0.75, respectively) are completely different. Furthermore, many respondents with a lower reuse ratio than respondent A have higher VI values. These results show how misleading the reuse ratio can be in diagnosing security vulnerability due to its inability to incorporate the skewness of the usage of login credentials.

Fig. 4 shows the distribution of the VI values depending on the number of unique (ID, PW) pairs: we observe a large variation of the VI values for a given m. For example, the respondent corresponding to point A has a low VI of 0.15 with $m = 10$. On the other hand, the two respondents corresponding to the points within the circle B have very high VI values of 0.75 and 0.69, respectively, with m values (10 and 11, respectively) that are similar to that of respondent A. This means that even with the same number of unique (ID, PW) pairs used, the vulnerability of users' login credentials tends to vary considerably, depending on how users allocate the pairs to the sites to which they subscribe.

Fig. 4 also shows that a smaller number of unique (ID, PW) pairs does not necessarily lead to a higher level of vulnerability. Let us consider point A again. While 27 respondents use a higher number of unique (ID, PW) pairs than respondent A, 22 of them have higher VI values than respondent A. In addition, let us compare respondents A and C. While respondent C uses a considerably higher number of unique (ID, PW) pairs ($m = 23$) than respondent A ($m = 10$), there is no significant difference between the vulnerability of their respective credentials.

These observations imply that the security of Internet users' login credentials can be significantly improved without creating new IDs, PWs, or (ID, PW) combinations, which may be a challenge because of users' cognitive limitations. To verify this, we calculate the gap between VI and VI_{min} for each respondent. Fig. 5 shows the distribution of the gaps over the number of unique (ID, PW) pairs. We find that the gaps are substantial, with an average of 0.29. By comparing with the average VI of 0.38 (Fig. 4), we can see that the respondents' VI values can be reduced by 76% (average) by uniformalizing the usage of (ID, PW) combinations. Note that the reuse ratio remains the same under uniformalization, again showing its limitation.

The measure of VI implicitly assumes that the potential loss caused by a security breach is uniform across all sites.[7] However, different sites can have different values for a user and therefore the potential losses can also vary. For example, sites that involve financial transactions, such as banking sites, or sensitive private information, such as SSNs and health care records, are usually more important than other sites. This variation can be incorporated easily by plugging the losses into the VI formula as follows:

$$\text{Revised VI} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\frac{1}{N}\right)\left(\frac{\sum_{l=1}^{n_i} w_{il} - w_{ij}}{\sum_{k=1}^{m}\sum_{l=1}^{n_k} w_{kl} - w_{ij}}\right), \qquad (2)$$

where w_{ij} denotes the loss from a security breach at the jth site in component i. Note that with $w_{ij} = 1$ for all i and j, Eq. (2) is reduced to Eq. (1).

To sum up, relying on a network perspective for login credentials vulnerability, the proposed measure of VI incorporates the structure of the (ID, PW) usage network by linking the vulnerability with the distribution of (ID, PW) combinations over multiple accounts. The application of the measure to the sample data shows that the reuse ratio can be significantly misleading about the vulnerability current Internet users face and that users can substantially reduce vulnerability by balancing their login credentials usage.

6. Discussion and conclusion

6.1. Why so vulnerable?

The reason why Internet users' behavioral patterns of ID and PW usage make their login credentials vulnerable can be explained using cybernetic theory and cognitive psychology theory. According to cybernetic theory, a discrepancy-enlarging feedback loop is involved in acts of avoidance, as in reducing security vulnerability (Carver & Scheier, 2002; Liang & Xue, 2009). This loop is triggered by identifying one's present state (e.g., present vulnerability) and comparing it with an undesired end state (e.g., being the victim of security breach). If both states are too close, a behavior is activated to make changes to enlarge the gap between them. These processes together form a discrepancy-enlarging feedback loop.

Anecdotal evidence from this study suggests that this feedback loop is unlikely to be effective in the case of most Internet users. More specifically, during the course of data gathering, we found that almost no respondents had any idea about the current state of vulnerability of their login credentials. When the data gathering was finished, many of the respondents were surprised at their behavioral patterns, especially the small numbers of PWs they were using (identification of the present state and determination of the closeness between the present state and the undesired state). Some

[7] The measure of the reuse ratio is also based on the same implicit assumption.

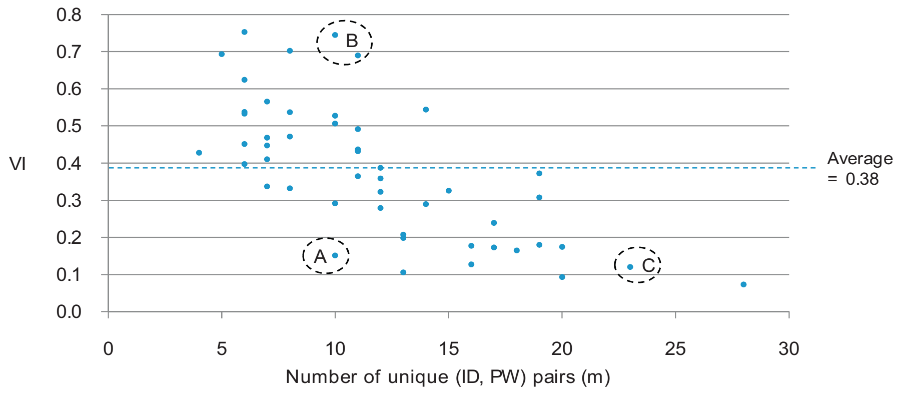


Fig. 4. VI values depending on the number of unique (ID, PW) pairs.

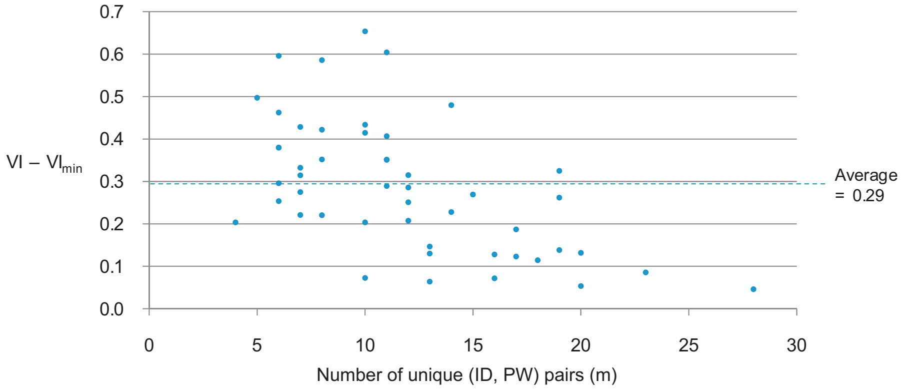


Fig. 5. Potential reduction in vulnerability.

of them said that they should use more PWs to reduce the security risks (intention to make changes). Therefore, users need to know their current state of vulnerability to trigger feedback loops for reducing vulnerability. For doing so, the VI can be an effective diagnostic measure. As a practical method, we recommend the Internet user apply and calculate the VI values for a subset of the sites he or she is subscribing to (e.g., 20–30 frequently used sites) instead of all of them.

Cognitive psychology theory implies that while people may be able to remember a few unique (ID, PW) combinations without difficulty, as the number of combinations increases, they have great trouble remembering them. As a result, a security–convenience tradeoff exists.

The supplementary data set supports the security–convenience tradeoff. Many websites request that their users periodically change their PWs. We surveyed the respondents' reactions to the request and found that 80% of users kept their current PWs when possible. In addition, 16% of the respondents said that they changed their current PW to one of the PWs they were using on another site. Only 4% of the respondents answered that they created a completely new PW.[8]

To enhance memory of Internet users, some mnemonic techniques can be applied to the traditional ID–PW based user-authentication mechanism in websites. Nelson and Vu (2010) showed that image-based mnemonic techniques can help users memorize and recall their PWs effectively, compared to cases in which proactive password checking restrictions or text-based mnemonic techniques are applied.

6.2. Implication

From a practical viewpoint, the results of this study suggest several recommendations to firms and policy makers addressing the issue of login credentials vulnerability. First, firms need to have a network perspective on the security of users' login credentials and be acquainted with their linkages with other firms in terms of security vulnerability. With this perspective, they are advised to collaborate with other firms. Given the network nature of login credentials and the accompanying vulnerability, as in the case of Naver.com mentioned previously, firms should understand that the efforts to improve the security of their own sites or systems are not satisfactory. Instead, in extreme cases, firms can improve their security more effectively by supporting the security improvement efforts of other firms with fewer resources rather than by focusing on their own security improvements. Major firms would want to lead the organization and funding of these collaborative efforts. The largest telecommunications service company in South Korea, is a good example. As an industry-wide collaborative effort, it provides security solutions to small and mid-sized Internet businesses (Kwon, 2010).

Second, firms would want to develop and implement new authentication systems other than IDs and PWs. Considering the inherent behavioral limitations of users, IDs and PWs are inherently vulnerable. In the long-term, new authentication systems that are less subject to the behavioral limitations of human beings should be implemented. This recommendation is equally applicable to government agencies. To complement the vulnerability of login credentials, public key certificate-based authentication mechanism

[8] Given the PW change request, some users may change their current PWs and keep an electronic list of their (site, ID, PW) combinations, which has its own significant security risks. We thank a reviewer for this insight.

has been widely adopted among online firms, especially for online banking and commerce sites. The adoption should be expanded to sites in other areas. Other possible mechanisms to adopt include image authentication (Chang & Lin, 2008; Renaud, 2009), two or multi-factor authentication with biometrics information (Apampa, Zhang, Wills, & Argles, 2008; Bhargav-Spantzel et al., 2007) and one-time password authentication based on time and users' location (Wen-Bin & Jenq-Shiou, 2011).

Third, policy makers must enforce the implementation of security measures for login credentials across the board. Many countries have been forcing firms to implement security measures in a selective manner; that is, some firms are subject to enforcement while others not. The South Korean government, for example, requires about 1000 major websites (portals with more than 50,000 visitors a day and websites with more than 10,000 visitors a day) to meet specific guidelines so that identity theft can be prevented. The Identity Theft Red Flags Rule in the US, issued in 2007, requires creditors and financial institutions to implement identity theft prevention programs. These guidelines require creditors and financial institutions with "covered accounts" to develop and employ written identity theft prevention programs (Finklea, 2010). However, network perspective analysis suggests that these policies may not be effective, even for the relevant institutions, if they are linked to other vulnerable sites or to institutions that are exempted from the mandatory implementation. Thus, we must focus on increasing the security level of medium and small organizations, which are often more vulnerable to identity theft.

Finally, the public awareness of security needs to be improved, as a general approach to facilitate vulnerability-reducing feedback loops. Specifically, awareness about not only overall identity security but also the management of login credentials based on the network perspective is required.

6.3. Conclusion

This study aimed to advance our knowledge of login credentials vulnerability on the Internet and to improve information security management practices for login credentials. On the basis of unique data from Internet users and a novel perspective on login credentials usage, this study made the following contributions.

First, while most existing studies have provided usage statistics of login credentials from recall-based survey data, this study is based on the actual data set on the usage. Our analysis contributes to the information security literature by showing that recall may not be credible and thus a recall-based study tends to generate a biased picture of login credentials usage, usually underestimating the vulnerability. Specifically, we find that the same login credentials are used for more accounts and reused more often than previously suggested in the literature.

Second, this study contributes to the security research by showing the limitations of current vulnerability measures of login credentials and by proposing a new vulnerability measure from a network perspective. Based on this perspective, we find that Internet users' login credentials usage patterns are significantly skewed. The most frequently used combination of ID and PW for each user is used for as many as 54% of all the sites to which the user subscribes. Meanwhile, the current vulnerability measures of login credentials either fail to consider the reuse of login credentials over multiple accounts (e.g., PW strength) or do not reflect the skewness of usage patterns (e.g., PW reuse ratio). By relying on a network perspective for login credentials vulnerability, we suggest a new vulnerability measure of individual users that captures the structural characteristics of the ID–PW usage network. The suggested measure VI can be used to enhance our understanding on login credential vulnerability by considering a behavioral pattern of the usage of login credentials, which is generally highly skewed.

Finally, this study contributes to the information security management practices by providing several implications for managers and policy makers striving to reduce security vulnerability.

There are three areas that warrant further research. First, the results of our study suggest that behavioral research on security needs to be more rigorous to ensure that accurate data are considered. Since the speculation-based data obtained from users may be unreliable, more objective data on users' behavior are an essential prerequisite for verifying the validity of research. Therefore, more research to investigate easy methods for obtaining reliable data on users' behaviors is needed. Second, while the suggested measure of the VI incorporates the structure of the login credentials usage network, it does not consider the characteristics of individual login credentials. Taking into account the strength and complexity of user PWs to upgrade the measure would be an important venue for further research. Third, determinants of Internet users' ID–PW usages patterns need to be studied further. During the investigation of Internet users' ID–PW usages, we found that the variation of VI values is large among respondents. Which factors influence the VI of an Internet user? Why do some people manage their login credentials better than others? These are other venues for further research.

Acknowledgment

This research was financially supported by Hansung University.

References

Adams, A., & Sasse, M. A. (1999). Users are not the enemy. *Communications of the ACM*, *42*(12), 41–46.

Alhazmi, O. H., Malaiya, Y. K., & Ray, I. (2007). Measuring, analyzing and predicting security vulnerabilities in software systems. *Computers & Security*, *26*(3), 219–228.

Anderson, C. L., & Agarwal, R. (2010). Practicing safe computing: A multimethod empirical examination of home computer user security behavioral intentions. *MIS Quarterly*, *34*(3), 613.

Apampa, K. M., Zhang, T., Wills, G. B., & Argles, D. (2008). Ensuring privacy of biometric factors in multi-factor authentication systems. In *International conference on security and cryptography in ICETE 08* Portugal, Porto,

Bhargav-Spantzel, A., Squicciarini, A. C., Modi, S., Young, M., Bertino, E., & Elliott, S. J. (2007). Privacy preserving multi-factor authentication with biometrics. *Journal of Computer Security*, *15*(5), 529–560.

Blumenberg, S., Wagner, H.-T., & Beimborn, D. (2009). Knowledge transfer processes in IT outsourcing relationships and their impact on shared knowledge and outsourcing performance. *International Journal of Information Management*, *29*(5), 342–352.

Brown, A. S., Bracken, E., Zoccoli, S., & Douglas, K. (2004). Generating and remembering passwords. *Applied Cognitive Psychology*, *18*(6), 641–651.

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*, *34*(3), 523–527.

Bunnell, J., Podd, J., Henderson, R., Napier, R., & Kennedy-Moffat, J. (1997). Cognitive, associative and conventional passwords: Recall and guessing rates. *Computers & Security*, *16*(7), 629–641.

Burr, W. E., Dodson, D. F., & Polk, W. T. (2006). Information security: Electronic authentication guideline. *NIST special report* (pp. 800–863).

Carver, C. S., & Scheier, M. F. (2002). Control processes and self-organization as complementary principles underlying behavior. *Personality and Social Psychology Review*, *6*(4), 304–315.

Chang, C. C., & Lin, P. Y. (2008). A color image authentication method using partitioned palette and morphological operations. *IEICE Transactions on Information and Systems*, *91*(1), 54–61.

Chang, R. M., Oh, W., Pinsonneault, A., & Kwon, D. (2010). A network perspective of digital competition in online advertising industries: A simulation-based approach. *Information Systems Research*, *21*(3), 571–593.

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. *Information Systems Research*, *20*(1), 79–98.

Doherty, N. F., Anastasakis, L., & Fulford, H. (2011). Reinforcing the security of corporate information resources: A critical review of the role of the acceptable use policy. *International Journal of Information Management*, *31*(3), 201–209.

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, *8*, 128–140.

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the Internet topology. *SIGCOMM Computer Communication Review*, *29*(4), 251–262.

Farahmand, F., Navathe, S. B., Sharp, G. P., & Enslow, P. H. (2005). A management perspective on risk of security threats to information systems. *Information Technology and Management*, *6*(2), 203–225.

Finklea, K. M. (2010). *Identity theft: Trends and issues*. DIANE Publishing Company.

Florencio, D., & Herley, C. (2007). A large-scale study of web password habits. In *Proceedings of the 16th international World Wide Web conference*. Banff, Alberta, Canada: ACM Press.

Gaw, S., & Felten, E. W. (2006). Password management strategies for online accounts. In *Symposium on usable privacy and security* Pittsburgh, PA.

Goldenberg, J., Libai, B., Muller, E., & Stremersch, S. (2010). Database submission—The evolving social network of marketing scholars. *Marketing Science*, *29*(3), 561–567.

Hansen, M. T. (2002). Knowledge networks: Explaining effective knowledge sharing in multiunit companies. *Organization Science*, *13*(3), 232–248.

Horcher, A. M., & Tejay, G. P. (2009). Building a better password: The role of cognitive load in information security training. In *IEEE international conference on intelligence and security informatics*.

Ives, B., Walsh, K. R., & Schneider, H. (2004). The domino effect of password reuse. *Communications of the ACM*, *47*(12), 75–78.

Janhonen, M., & Johanson, J.-E. (2011). Role of knowledge conversion and social networks in team performance. *International Journal of Information Management*, *31*(3), 217–225.

Javelin Strategy & Research. (2010). *Javelin study finds identity fraud reached new high in 2009, but consumers are fighting back*. https://www.javelinstrategy.com/news/831/92/Javelin-Study-Finds-Identity-Fraud-Reached-New-High-in-2009-but-Consumers-are-Fighting-Back/d, pressRoomDetail

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabasi, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651–654.

Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: An empirical study. *MIS Quarterly*, *34*(3), 549–566.

Kane, G. C., & Alavi, M. (2008). Casting the net: A multimodal network perspective on user–system interactions. *Information Systems Research*, *19*(3), 253–272.

Kankanhalli, A., Teo, H.-H., Tan, B. C. Y., & Wei, K.-K. (2003). An integrative study of information systems security effectiveness. *International Journal of Information Management*, *23*(2), 139–154.

Kaspersky Lab. (2007). *Online accounts vulnerable to identity theft, says Kaspersky Lab*. http://www.kaspersky.com/about/news/press/2007/Online_Accounts_Vulnerable_to_Identity_Theft_says_Kaspersky_Lab

Korea Communications Commission and Korea Internet & Security Agency. (2009). *2009 survey on the Internet usage: Executive summary*. http://isis.kisa.or.kr/board/index.jsp?pageId=040100&bbsId=7&itemId=728&pageIndex=2

Korea Communications Commission and Korea Internet & Security Agency. (2010). *2010 survey on the Internet usage: Executive summary*. http://isis.kisa.or.kr/board/index.jsp?pageId=040100&bbsId=7&itemId=774&pageIndex=1

Kwon, C. (2010). KT providing free security solution. *Computer Times*, http://www.computertimes.co.kr/news/articleView.html

LaRose, R., Rifon, N. J., & Enbody, R. (2008). Promoting personal responsibility for internet safety. *Communications of the ACM*, *51*(3), 71–76.

Liang, H. G., & Xue, Y. J. (2009). Avoidance of information technology threats: A theoretical perspective. *MIS Quarterly*, *33*(1), 71–90.

McMillan, R. (2011). Sony cuts off Sony online entertainment service after hack. *Computer World*, http://www.computerworld.com/s/article/9216343/Sony_cuts_off_Sony_Online_Entertainment_service_after_hack

Miller, G. A. (1994). The magical number 7, plus or minus 2—Some limits on our capacity for processing information(reprinted from psychological review, vol. 63, pg. 81, 1956). *Psychological Review*, *101*(2), 343–352.

Miniwatts Marketing Group. (2011). *Internet world stats: Usage and population statistics*. http://www.internetworldstats.com

National Internet Development Agency of Korea. (2009). *Internet usage comparison between Korea and the U.S.* http://isis.kisa.or.kr/board/index.jsp?pageId=040100&bbsId=7&itemId=768&pageIndex=2

Nelson, D., & Vu, K. P. L. (2010). Effectiveness of image-based mnemonic techniques for enhancing the memorability and security of user-generated passwords. *Computers in Human Behavior*, *26*(4), 705–715.

Nooy, W. D., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. New York: Cambridge University Press.

Patel, S. C., Graham, J. H., & Ralston, P. A. S. (2008). Quantitatively assessing the vulnerability of critical information systems: A new method for evaluating security enhancements. *International Journal of Information Management*, *28*(6), 483–491.

Pernul, G. (1995). Information systems security: Scope, state-of-the-art, and evaluation of techniques. *International Journal of Information Management*, *15*(3), 165–180.

Rapoport, A., & Horvath, W. J. (1961). A study of a large sociogram. *Behavioral Science*, *6*(4), 279–291.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B: Condensed Matter and Complex Systems*, *4*(2), 131–134.

Renaud, K. V. (2009). Guidelines for designing graphical authentication mechanism interfaces. *International Journal of Information and Computer Security*, *3*(1), 60–85.

RSA. (2004). *RSA security study shows identity theft awareness high, but consumer confidence low*. http://www.rsa.com/press_release.aspx?id=3377

Schneier, B. (2000). *Secrets & lies: Digital security in a networked world*. New York: Wiley Computer Publishing.

Siponen, M., & Vance, A. (2010). Neurtralization: New insights into the problem of employee information systems security policy violations. *MIS Quarterly*, *34*(3), 487.

Straub, D. W. (1990). Effective IS security: An empirical study. *Information Systems Research*, *1*(3), 255–276.

Straub, D. W., & Welke, R. J. (1998). Coping with systems risk: Security planning models for management decision making. *MIS Quarterly*, *22*(4), 441–469.

Syson, F., & Perks, H. (2004). New service development: A network perspective. *Journal of Services Marketing*, *18*(4-5), 255–266.

The Nielsen Company. (2010a). *Top online sites and brands in the U.S.* http://blog.nielsen.com/nielsenwire/online_mobile/june-2010-top-online-sites-and-brands-in-the-u-s

The Nielsen Company. (2010b). *Top U.S. web brands and site usage*. http://blog.nielsen.com/nielsenwire/online_mobile/top-u-s-web-brands-and-site-usage-december-2009

Vu, K.-P. L., Proctor, R. W., Bhargav-Spantzel, A., Tai, B.-L., Cook, J., & Eugene Schultz, E. (2007). Improving password security and memorability to protect personal and organizational information. *International Journal of Human-Computer Studies*, *65*(8), 744–757.

Wagner, A., & Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1478), 1803–1810.

Wang, N. C., Wang, C., & Wulf, W. A. (1997). Towards a framework for security measurement. In *20th national information systems security conference* Baltimore, MD, (pp. 522–533).

Weber, J. E., Guster, D., & Safonov, P. (2008). A developmental perspective on weak passwords and password security. *Journal of Information Technology Management*, *19*(3), 1–8.

Wen-Bin, H., & Jenq-Shiou, L. (2011). Design of a time and location based One-Time Password authentication scheme. In *Wireless communications and mobile computing conference (IWCMC)* Istanbul, Turkey.

Workman, M., Bommer, W. H., & Straub, D. (2008). Security lapses and the omission of information security measures: A threat control model and empirical test. *Computers in Human Behavior*, *24*(6), 2799–2816.

Yan, J., Blackwell, A., Anderson, R., & Grant, A. (2004). Password memorability and security: Empirical results. *IEEE Security & Privacy*, *2*(5), 25–31.

Zhang, J., Luo, X., Akkaladevi, S., & Ziegelmayer, J. (2009). Improving multiple-password recall: An empirical study. *European Journal of Information Systems*, *18*(2), 165–176.

Youngsok Bang is a postdoctoral fellow at McGill University. He received his BS, MS, and Ph.D. degrees in Management Engineering from KAIST. His current research interests focus on information systems economics and online security and privacy. His work has appeared in *MIS Quarterly*.

Dong-Joo Lee is an assistant professor at the Division of Management, Hansung University, in Seoul, Korea. He holds a Ph.D. in Management Engineering from the Graduate School of Management, KAIST. His research interests include information security and privacy, personalization, and information systems economics. His work has appeared in several journals, including *MIS Quarterly*, *European Journal of Operational Research*, *Long Range Planning*, *Technovation*, and *Knowledge Management Research and Practice*.

Yoon-Soo Bae is a doctoral candidate at the Graduate School of Management, KAIST. He received both his BS and MS degrees in Management Engineering from KAIST. His current research interests focus on consumer searching behavior and neuromarketing.

Jae-Hyeon Ahn is a professor at KAIST Business School in Seoul, Korea. He received both his BS and MS degrees from Seoul National University, Seoul, Korea, in 1984 and 1986, respectively, and his Ph.D. degree in decision sciences from Stanford University in 1993. After graduation, he worked as a senior researcher at AT&T Bell Labs from 1993 to 1998. His current research interests are focused on, among other things, investment strategies for information system security, neuro-marketing approaches for Internet business, and behavioral decision making. He has published papers in *MIS Quarterly*, *Management Science*, *Decision Support Systems*, and *Journal of Information Technology*, among others.