International Journal of Information & Network Security (IJINS)
Vol.2, No.1, February 2013, pp. 32~42
ISSN: 2089-3299

Identifying Phishing Threats in Government Web Services

Yunsang Oh*, Takashi Obi**
* Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology
** Imaging Science and Engineering Laboratory, Tokyo Institute of Technology

Article Info

Article history:

Received July 20th, 2012
Accepted Aug 28th, 2012

Keyword:

E-Government
Phishing
Privacy

ABSTRACT

The governmental use of Web technologies, including e-Government, has many advantages for citizens, but progress in this relationship has highlighted information security as an important issue in preserving a citizen's privacy. Unfortunately, unique governmental characteristics lead users to authenticate its service unwillingly; users may investigate service's possible and likely vulnerabilities carelessly when perceiving trustworthiness. In this paper, we study a threat model about how government Web services become privacy leak targets, especially through phishing attacks. We identify three service characteristics, sensitivity, involuntarity, and linkability, and illustrate how phishers can effectively exploit these characteristics. Furthermore, we conducted a real phishing attack experiment, hijacking a government-certified commercial service in South Korea to complete our investigation. Finally, we propose mitigation strategies for building a trustworthy government Web service against phishing attacks.

Corresponding Author:

Yunsang Oh,
Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology,
4259-R2-60 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan
Email: oh.y.ab@m.titech.ac.jp

1. INTRODUCTION

The Internet is completely revolutionizing the relationship between the government, public sector au- thorities and citizens. Governmental use of Internet, such as e-Government, has many direct advantages for citizens and businesses [3]; however, progress in this relationship has highlighted information security as an important issue [20]. Governmental use of information security can raise the public's trust and confidence in online transactions. Many governments must therefore augment their use of security technologies for successful services because profound security risks and vulnerabilities must be managed.

Effectively secure government services require proactive IT governance and a robust infrastructure so that the services can handle electronic administrative transactions, including those in banking, employment, education, military service, taxation, and healthcare, which are considered highly private. For instance, in e-healthcare, the most promising personalized e-Government service (In a user survey targeting European citizens [2], 44% of the total sample, 71% of the Internet users, had used the Internet for health purposes. In another survey targeting Americans [8], 80% listed accessing medical information from the National Institute of Health as their most favored example of e-Government.), users seeking electronic medical information ranked personal privacy as their most important concern [19].

The success and acceptance of Web services are generally contingent upon users' willingness to adopt the services. This willingness is closely related to perceiving trustworthiness, which depends on users' confidence in the services and enabling technology [7]. Bélanger and Carter [4] noted that trusting the Internet is an essential element for government Web service adoption. Users must believe that mechanisms are in place to ensure secure and private data transmission over an impersonal medium. These authors pointed out that government agencies should first emphasize their general competence in their areas of expertise and then highlight their ability to provide services via the Internet.

Unlike commercial services' competition for survival, which weeds out incompetent players, government Web services are typically provided unilaterally without relevant user security evaluations. Governmental policies may force services upon users, even without users' willingness to adopt. In this paper, we point out that unique governmental characteristics can promote negative user behaviors, especially carelessness in investigating possible vulnerabilities when perceiving Web service trustworthiness. More specifically, we are concerned that government Web services form a new trend of online phishing. This deceit-based attack threatens the services' success because it erodes trust in the underlying infrastructure. We therefore identify possible security breaches in government-driven public Web services and propose defense strategies to mitigate phishing threats that could hijack services.

This paper mainly contributes an identification of a phishing threat model for hijacking government Web services (section 2). We conduct an empirical case study of the identity theft alert service in South Korea to identify potential phishing attack strategies exploiting government Web services' characteristics (section 3). The resulting real phishing experiment analysis completes our study (section 4). Last, as a subsidiary contribution, we review defense strategies for developing government Web services to mitigate phishing threats (section 5).

2. Threat Model

Phishing is a form of deception in which an attacker attempts to fraudulently acquire sensitive information from a victim by impersonating a trustworthy entity. For example, an attacker would gain the victims' trust using Context-Aware Phishing [13] where victims believe the received phishing messages' authenticity.

We are concerned with online phishing attacks impersonating the government. To examine the phishing threat model, we characterize a government Web service and develop a set of hypotheses concerning the phishers' malicious strategies that exploit these characteristics and deceive general users. We assume that governmental administrative entities typically deliver sensitive information to citizens, including legal issues, taxation, public security or military duties. The information is characterized by its exclusivity, as only these administrative entities can provide it. Users often unilaterally receive electronic messages containing this information. In this case, users usually anticipate damages, such as penalties for neglecting duties if they do not respond promptly to the requested actions. In the authoritative atmosphere, repeated interactions with the administrative transactions effectively train users to involuntarily authenticate such transactions without careful investigation. Administrative transactions require identification data issued to citizens and legal residents, allowing services to bind transactions to a user's real identity. Some countries use personal information such as the date or place of birth as a national identification number. The basic information can, however, sometimes be easily taken or inferred from open data repositories such as social network services. Personal data can be exploited as a quasi-identifier to link to a user's real identity. From these assumptions, we hypothesize that the phisher's strategy exploits the following characteristics:

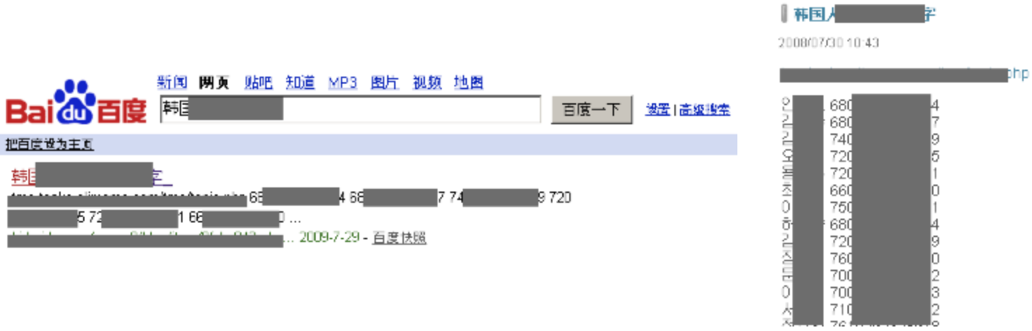
Sensitivity. A government unilaterally delivering highly sensitive information to users can be exploited to cheat victims because victims may either be seriously concerned with any disadvantage, such as penalties for neglecting duties, or overestimate sensitivity due to the information source's uniqueness and authority in the country.
Involuntarity. Government Web service authenticity can be effectively faked because citizens are forced and trained to authenticate administrative entities without relevant vulnerability investigation due to unilateral interactions.
Linkability. Context-aware phishing can be a new trend of impersonating government Web services, as phishers can effectively infer recipient's identification data and exploit it as bait. Data crawling attempts can target open data repositories, such as social networks, to reconstruct trustworthy identification data on a major scale.

In sections 3 and 4, our experimental case study suggests that the phishers' strategies are valid. Sections 3 and 4 are major revisions of our previous work [22].

3. Case Study of Identity Theft Alert Service in South Korea

3.1. Background

Many countries' governments use national identification numbers to track their citizens, residents, taxation, healthcare, and other administrative functions. In South Korea, a resident registration number also identifies people in most private and administrative transactions. A resident registration number is a 13-digit number that include digits about the person's date of birth, gender, and birth place. Interestingly, a user must submit her or his real name and resident registration number for identity validation to create Web service accounts in South Korea. Unfortunately, compulsory Internet real-name system threatens privacy, especially identity theft through hacking incidents. In our Web search conducted in March 2010, we easily found names and resident registration numbers of a few Koreans in a Web search engine. In the Chinese Web portal Baidu.com, we submitted 7 Chinese characters, meaning "Korean Identification Number", and found a URL in the first search result revealing numbers similar to Korean resident regis- tration numbers, as in Figure 1 (a). Clicking the link, we found names and resident registration numbers of 196 Koreans, as shown in Figure 1 (b).



韩国[redacted]

韩国[redacted]字

[redacted] 6[redacted]4 6[redacted]7 74[redacted]9 720 [redacted]5 72[redacted]1 66[redacted]0 ...

[redacted]... 2009-7-29 - 百度快照

韩国人[redacted]字

2008/07/30 10:43

[redacted]php

[redacted]

(a) Baidu.com Search (b) Search Result

Figure 1: Exposed Korean Identity Numbers in a Portal Site - Baidu.com Search

To lower the chances of forgery and privacy invasion that follow stolen resident registration numbers, the South Korean government introduced an alternative identifier that replaces a users' real name and resident registration number when signing onto Internet services. They also certified 5 credit information providers as issuers of an alternative identifier, called an i-Pin. The credit information providers provide services not only for i-Pin issuance, but also identity theft prevention. The providers issue an i-Pin to individuals for free, and some profit from a service that monitor resident registration numbers or i-Pin usage to alert or block suspicious identity validation attempts. If they detect an identity validation attempt using a resident registration number or i-Pin, the identifier's owner receives an alert by email or mobile phone message.

The identity theft alert contains several usability issues. An "email" alert from a governmental entity is highly vulnerable to phishing attacks impersonating the email sender. In this section, we empirically analyze whether this concern is valid using an actual (harmless) phishing experiment. To measure the success rate of a phishing attack that impersonates a government's service, we designed a field experiment mimicking a real phishing attack. Our experiment aims to deceive and lure target subjects into opening our email alert about a suspicious identity validation attempt and accessing our phishing Web site, which requests personal information about a real name and resident registration number.

Although the credit information providers are included in the private sector, the amount and sensitivity of the data they collect about citizens (such as real name, national identification number, financial activities, and personal credit rating) is significant enough that these providers are considered as one of the administrative entities. The government also certified their services and requested that Web service providers in South Korea adopt their services according to national law. We thus assume that impersonating a credit information provider produces a similar effect as impersonating a governmental entity.

3.2. Target Subjects

We randomly targeted South Koreans, selected from friends, colleagues, and anonymous people. To confirm that subjects regularly read their email, we sent an email request to join a survey without revealing our research theme. We selected 52 subjects who replied positively. Before our experiment, we did not notify the subjects about our phishing experiment.

The subjects included 37 males and 15 females, aged 29 to 67. We included 20 information technology experts in the subject group. The experts had at least one year of research or job experience related to cryptography, information security, or contents protection. Some subjects had already used similar services to monitor and receive alerts from the providers about suspicious identifier usages or financial transactions.

Because a deceit-based field study without subjects' prior knowledge of or consent to the experiment is ethically problematic, we apologized to all subjects when our experiment ended and received their consent to use the results for our research. All subjects' information was neither stored nor transferred over a network. Our study environment, such as the number of subjects, is inevitably limited in order to avoid all ethical and legal troubles.

3.3. Our Attack Strategies

We played the role of a phisher sending a phishing email from the fake (not existing) credit information service company "Seoul Credit Info Telecommunication [seoul- shin-yong-jung-bo-tong-shin]". We based our phishing strategy on the following assumptions. First, a phishing email with sensitive content such as a suspicious identity validation attempt would easily concern users, and an unexpected alert would effectively maximize subjects' concern. We assumed that subjects, worried about the email's contents, would reluctantly authenticate the email sender. Second, a registered user of a legitimate service provider would be familiar with the alert, but this familiarity would ironically cause carelessness when authenticating the service provider. Because the national identification data hacking incidents are common concerns shared throughout society, we can design a single phishing message that can be equally reused regardless of users' service experiences. Third, phishers would easily gain a user's trust if the phishing message included identification data, including real name, Web service login name or resident registration number. We assumed that users would unconsciously link these data to their real identities. In section 3.4, we describe generating the identification data.

3.4. Reconstructing a User's Identifier

A user's birth date and gender can produce a numeric string likely to be a resident registration number. We thus study how users perceive the reconstructed identification number. In practice, the personal data can be collected from social network services such as Facebook (www.facebook.com). Jagatic et al. [12] discussed honing phishing attacks with publicly available personal information from social networks. Unfortunately, Bonneau and Preibusch [6] noted that social network services do not effectively convey privacy control functions to users. They pointed out that, despite evidence that social network service providers are attempting to implement privacy-enhancing technologies, privacy is rarely a selling point and services have failed to promote existing privacy controls. Preventing adversaries from compiling significant user data is a major challenge for social network service providers [5].

Table 1: Survey for South Koreans on Internet Identifiers

Q1. Do you have user accounts in more than one Internet service?		
Yes	96.15%	(125)
No	0.00%	(0)
No Answer	3.85%	(5)
Q2. If you have more than one user account, do you always use a single user ID?		
Always use a single ID	9.23%	(12)
Try to use a single ID, but a trivial modification is allowed	46.92%	(61)
Use 2 ~ 3 IDs	36.92%	(48)
Use 4 ~ 5 IDs	3.85%	(5)
Use more than 6 IDs	0.77%	(1)
Definitely use a new ID for each Internet service	0.00%	(0)
No Answer	2.31%	(3)
Q3. If you use 3 or less user IDs, why? (Multiple answers)		
Easy to remember	68.46%	(89)
Well known to my family, friends, colleagues as my nickname	14.62%	(19)
Habitual choice	17.69%	(23)
Used them for a long time	26.92%	(35)
Others	3.08%	(4)
No Answer	10.00%	(13)

We also investigated whether a web service login name could also effectively work as bait to cheat victims. As Internet users likely use a single ID string for multiple Web services, we assume that users are trained to perceive a login name as an alternative identifier equivalent to a real name. Our simple anonymous online survey, as in Table 1, shows how users decide their Internet identifiers. In June 2010, we invited our friends, colleagues, and several anonymous South Koreans, for a total of 130 participants. All participants answered that they have user accounts on more than one Internet service (Q1). As assumed, 46.92% of participants use a single identifier with a trivial modification (Q2) and 9.23% use a single identifier. In addition, 36.92% use 2 or 3 identifiers. This result shows that 3 or fewer identifiers can locate 93.07% of participants: participants have used these identifiers for a long time and remember them easily (26.92% and 68.46% in Q3, respectively). Because a login name is generally used as part of an email address in most commercial services, we took the first word of an email address as the web service login name. Our strategy succeeded if the phishing victims perceived the login name as an authentic identifier.

3.5. Experiment Procedure

We conducted this experiment from May to June 2010. We divided subjects into 4 groups depending on the recipient name field in the phishing email. As shown in Figure 2 (a), Group 1 received an email with the recipient’s real name in the Korean alphabet. Group 2 received an email with a name in Korean and an inferred resident registration number. Groups 3 and 4 received an email with the first word in an email address instead of a real name. As with Group 2, an email for Group 4 also contained an inferred resident registration number. For a resident registration number, only 6 (the date of birth field) of the full 13 digits were visible. As mentioned in section 3.3, we inferred the number from the date of birth data found in social network websites.

As shown in Figure 2 (b), we made a phishing email alerting users that someone else used their identifiers. We added a date-time field to show when their identifier was used. The email also included a link to our fraudulent site requesting subjects to submit their names and resident registration numbers to find details of this suspicious identifier usage. If a subject submitted the identity information, she or he saw our apology and experiment explanation. We imitated the “Look and Feel” of a notification email from an existing credit information providing company. The phishing email and fraudulent site included PHP codes to record the access time of each step.

3.6. Results

As shown in Table 2, we observed that 35 of 52 subjects (67.3%) read the phishing email, and 29 (55.8%) clicked to access the fraudulent site asking for a name and resident registration number. Finally, 13 subjects (25%) submitted their names and resident registration numbers. We observed that 29 of 35 subjects (82.86%) who opened the email alert clicked the link to our phishing site. This result clearly demonstrates that our strategy for generating email contents successfully lured subjects.

Table 2: Experiment Results

	Target	Open Phishing Email	Access Phishing Site	Submit Info
All	52	35 (67.3%)	29 (55.8%)	13 (25.0%)
Group 1	14	9 (64.3%)	7 (50.0%)	3 (21.4%)
Group 2	12	7 (58.3%)	6 (50.0%)	3 (25.0%)
Group 3	13	9 (69.0%)	8 (61.5%)	5 (38.5%)
Group 4	13	10 (76.9%)	8 (61.5%)	2 (15.4%)

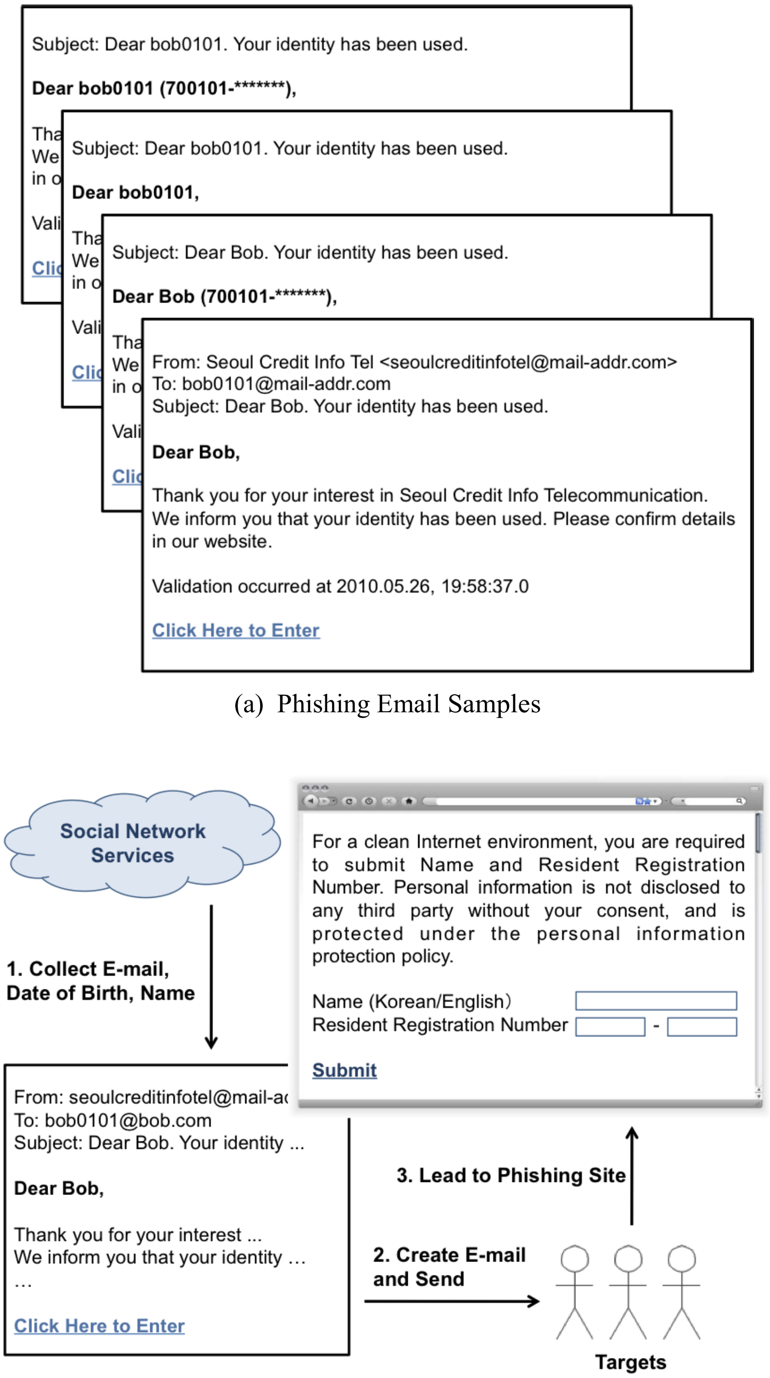
Table 3: Experiment Results for the IT Expert Group

	Target	Open Phishing Email	Access Phishing Site	Submit Info
IT Experts	20	14 (56.0%)	10 (50.0%)	5 (25.0%)

Table 4: Experiment Results based on Gender

	Target	Open Phishing Email	Access Phishing Site	Submit Info
Female	15	10 (66.7%)	8 (53.3%)	5 (33.3%)
Male	37	25 (67.6%)	21 (56.8%)	8 (21.6%)

Figure 2: Phishing Experiment



Subject: Dear bob0101. Your identity has been used.

Dear bob0101 (700101-*****),**

Subject: Dear bob0101. Your identity has been used.

Dear bob0101,

Subject: Dear Bob. Your identity has been used.

Dear Bob (700101-*****),**

From: Seoul Credit Info Tel <seoulcreditinfotel@mail-addr.com>
To: bob0101@mail-addr.com
Subject: Dear Bob. Your identity has been used.

Dear Bob,

Thank you for your interest in Seoul Credit Info Telecommunication. We inform you that your identity has been used. Please confirm details in our website.

Validation occurred at 2010.05.26, 19:58:37.0

Click Here to Enter

(a) Phishing Email Samples

For a clean Internet environment, you are required to submit Name and Resident Registration Number. Personal information is not disclosed to any third party without your consent, and is protected under the personal information protection policy.

Name (Korean/English) □
Resident Registration Number □ - □

Submit

From: seoulcreditinfotel@mail-ac
To: bob0101@bob.com
Subject: Dear Bob. Your identity ...

Dear Bob,

Thank you for your interest ...
We inform you that your identity …
…

Click Here to Enter

Social Network Services
1. Collect E-mail, Date of Birth, Name
2. Create E-mail and Send
3. Lead to Phishing Site
Targets

(b) Phishing Flow

Table 2 shows a similar trend in all groups: people are equally likely to authenticate all identifier types in the phishing message. Table 3 shows that the IT expert group did not show a different trend. Analyzing gender in Table 4, females and males also have a similar trend. Females were slightly more likely to become identity theft victims: 33.3% versus 21.6% for males.

4. Result Analysis

Our key finding is that phishing attacks impersonating a credit information provider definitely work. Based on the experimental result, we examine how the characteristics of government Web services - sensitivity, involuntarity, and linkability - influenced subjects' behaviors. The following analysis shows that a phishing strategy exploiting these three characteristics effectively cheats victims.

4.1. Subjects' Behavior Study

4.1.1. Sensitivity Influence

Our attack strategy succeeded if subjects perceived our message's sensitivity. To study subjects' perceptions about the phishing email's contents, we analyzed the promptness of their behavior. We assumed that the phishing email read time was inversely proportional to the victims' acceptance of the email's seriousness and authenticity. Specifically, if the email's contents influenced subjects at an early attack stage, their promptness to proceed would differ from uninfluenced subjects'.

Stay Duration at Phishing Email Of the 29 subjects who opened the fraudulent site, we successfully measured the email read time (from opening the email to clicking the link to the phishing site) of 23 subjects. Surprisingly, 20 subjects stayed at the email for less than 1 minute (18 seconds average). To study the correlation between the subjects' initial perception and later behaviors, we divided the 20 subjects into 2 groups: a "deceived" group of 10 subjects who submitted their personal information to our phishing site and a "robust" group of 10 subjects who refused to submit their information, as shown in Table 5 (The other 3 subjects of 23 stayed at the email for 2 minutes 1 second (in Deceived Group), 10 hours 9 min 5 seconds (in Robust Group), and 11 hours 56 min 1 second (in Robust Group). We ignore the three subjects in the consideration because of their values' significant distance from the others.). Interestingly, the "deceived" group stayed on the email for an average of 12.8 seconds, whereas the "robust" group stayed for 23.2 seconds. In the t-test, the email read time difference between the two groups is significant ($p \leq 0.02$). Through observation, we found that the phishing email contents in the early stage were sensitive enough to deceive subjects in the "deceived" group, and this initial perception influenced their behaviors in the later stages.

Stay Duration at Phishing Site Of the 13 subjects who submitted a name and resident registration number, 10 subjects stayed at the fraudulent site for less than 1 minute (19 seconds on average). This result implies the subjects mostly trusted the email's contents and did not hesitate to submit their private data. The other 3 subjects stayed at the fraudulent site for 1 min 36 seconds, 3 min 28 seconds, and 50 min 11 seconds. The last subject, even though he confirmed that his identification data had not been exploited at a legitimate service, inevitably submitted his data because of concerns about the email alert.

Table 5: Email read time of 20 subjects stayed in the email less than 1 minute (Unit: Second)

"Deceived" Group	4	6	7	7	7	13	14	18	24	28
"Robust" Group	9	15	17	19	19	22	26	30	36	39

4.1.2. Involuntarity Influence

To study whether the subjects' identity theft prevention service experiences influenced their behavior, we interviewed 17 of the 29 subjects who opened the phishing site from the email, and 10 subjects answered they had used a similar service. The visual deception imitating "Look and Feel" worked effectively: 9 of these 10 subjects answered that their experiences definitely influenced their behavior because they confused the service provider's name or "Look and Feel" of the email.

Interestingly, 7 subjects answered that they were deceived though they had no experiences with such services. One interviewee answered that he could not ignore our email because he once became a victim of a phishing attempt wherein a phisher impersonated the interviewee through an instant messenger and asked his friend to remit money. Another interviewee answered that he was concerned about the email because he had read some news articles about identity theft. These statements demonstrate that personal experiences or socially shared concerns can lead victims to involuntarily authenticate phishing messages, even without service experiences.

From these results, we studied two interesting behaviors: image activation in HTML-formatted email and browser plug-in installation. Both involuntary behaviors are equally vulnerable in security usability, but subjects behaved differently with interface authentication, depending on their familiarity with the risk. South Korean Internet users have recently faced debates through mass media about the usability of and potential security risks in ActiveX technology for Internet banking [15]. Conversely, the media has rarely mentioned image activation vulnerability, described above. Our observation below shows that the users' decisions in image activation and browser plug-in were different. This finding implies that user education through mass media is a likely defense strategy and worth implementing.

Image Activation To record email access time, we linked a PHP code to an <img> (image) HTML tag in the email. The PHP code is activated when images in the email are enabled. In the experiment, we successfully acquired the email access time of 29 subjects using the PHP code. Only 6 subjects of the 35 who opened the phishing email did not enable images. Most subjects answered that they habitually activate images in HTML-formatted emails. We also found that some Korean email service providers did not block images for HTML-formatted emails by default. Activating images in HTML-formatted pages from unauthenticated entities creates vulnerabilities, as the <img> tag is an often overlooked but convenient means for a Cross Site Scripting (XSS) attack [11]. The attacker can inject script contents into an image tag to steal information from a victim's browser and execute malicious scripts.

Browser Plug-In Refusal We observed subjects' reactions when requested to install a browser plug-in program. Korean Internet banking requires installation of the ActiveX [21] browser plug-in to protect transactions in integrity, confidentiality, and non-repudiation. Kim et al. [14] demonstrated that the ActiveX technology's popularity is a unique trend among banking systems in South Korea, but it harms Web standard compatibility and Web experience usability without enhancing security. Credit information providers also request that users install the ActiveX plug-in to properly use their services.

This uncomfortable experience influences subjects' reactions to the phishing email. For 14 of the 29 subjects who accessed the fraudulent site, we exposed a plug-in installation message using a widely used online banking plug-in name and "YES" / "NO" buttons. Of these 14 subjects, 12 refused to install a plug-in, and 1 closed the browser without making a selection; only 1 opted to install the plug-in. The subjects' average selection time was 5 seconds. Interestingly, only 3 of the 14 subjects who were requested to install the plug-in submitted their name and resident registration number, while 10 of the 15 subjects who were not asked to install the plug-in submitted their private data. Here, we found that subjects were deterred by browser plug-in programs and refused installation without hesitation.

4.1.3. Linkability Influence

To investigate the influence of various types of recipient's identification data, we compared the trends of 4 subject groups. Subjects in all groups were equally fooled by all types of recipient's identification data inferred from the email address or date of birth, i.e., all types of identification data worked equally effectively. Even with a simple ID taken from an email address without a real name, we effectively cheated subjects.

Identification Number Influence For 25 of all 52 subjects, our phishing email included a resident registration number whose last 7 digits were hidden with ∗, as in "700101-∗∗∗∗∗∗∗". For example, the visible 6 digits come from the date of birth for a person born on Jan. 1, 1970. This obfuscating expression is generally used in South Korea for privacy when publishing the registration number. We interviewed 8 subjects who read the email with their inferred resident registration number, and 6 answered that they intuitively recognized the numeric string as a legitimate identification number. Some subjects asked how we obtained their resident registration number and were surprised that they were fooled by the date of birth data only. They believed that the email sender knew every digit of their resident registration number. This result implies that a numeric string based only on the date of birth can easily masquerade as a real national identification number and can be exploited to attack South Korean Internet users.

Email Login Name Influence To study the influence of the identifiers taken from email addresses, we interviewed 8 subjects in Groups 3 and 4. Of them, 5 interviewees answered that they thought the email sender knew their exact name. This confusion occurred because the subjects used a string generated from their real names or nicknames for their login names. They had no doubt about the identifier in the recipient field though it was written in English, not Korean. In South Korea, English is rarely used in official communications.

5. Mitigation Strategy

We have thus far demonstrated that phishing can exploit governmental characteristics to cheat users. Having identified the phishing threat model targeting government Web services, we now consider improving overall security and usability. Providing effective Web browser interfaces that show and explain threat warnings is a usability challenge for preventing phishing attacks. There are already outstanding works that find effective anti-phishing tools; however, no work provides a complete solution. Combining worthy implementation proposals with proactive information security governance on a nationwide scale can enhance safety from phishing. Endless user education effectively complements this strategy. In this section, we propose defense strategies to thwart online phishing attacks that hijack the government. Each strategy has drawbacks, so we also discuss directions for future work.

5.1. User Education

User education is a key anti-phishing solution in the sensitivity and involuntarity perspectives. Providing education materials about phishing attacks was proved an effective method [17]; materials can present these concepts as comic scripts - defining phishing, steps to follow to avoid phishing attacks, and how criminals conduct phishing attacks. Online games also effectively teach users how to avoid phishing attacks [25]. However, emails with sensitive contents from authoritative sources still cause cognitive biases in overestimating sensitivity and bypass knowledge and experience when detecting evidence of Web page vulnerabilities. Governments should, in their communication principles, inform users that they avoid unnecessary personal information sharing in electronic messages like emails. In practice, user education success depends on effectively delivering education media and target audience sizes. As we observed in section 4.1.2, implementing user education through mass media is a likely phishing defense strategy.

From the linkability perspective, users must learn to be careful when providing basic personal informa- tion to private Internet service providers, such as social networks. Users can preserve privacy by properly using privacy control functions implemented by the providers, so no personal information is extracted from the personally defined boundary. However, some issues remain unresolved: first, we lack confidence in privacy control functions' effectiveness; second, we cannot assume that service providers themselves are trustworthy. Privacy control issues in private Internet services should be carefully studied, especially when the government uses personal information for the citizens and residents identification system.

5.2. Single Authentication Page

To prevent the most common phishing attempt, taking service login credentials (i.e., a login name and password) at the technical level, we recommend that government Web service users submit credentials to a single trusted authentication page rather than separate login pages of governmental entities. In practice, however, the authentication gateway itself can also be the target of phishing attack. Freedom in authentication gateway usage results in better authentication solutions. This freedom can provide users with more options in authentication methods, including an SSL certificate, one-time password, and smart card. A sign-in seal also helps users ensure that they are on a legitimate site, but a site authentication image does not always guarantee that a connection is secure or that it is safe to enter a password [24].

Despite these drawbacks, maintaining a single trusted authentication page increases the cost-effectiveness of anti-phishing countermeasures and reduces the server-side overhead of implementing anti-phishing tools for each service. From the sensitivity perspective, users can confirm sensitive messages from the admin- istrative entities not in the email but in the trusted page after login. Researchers in Japan have recently proposed and tested the scheme of providing a single user account for multiple governmental services as a reliable contact point [16].

5.3. Inference Attack Defense

From the linkability perspective, user de-identification in Web services reduces the probability of linking between online data and real identity. At first glance, it provides reasonable network anonymity. In practice, however, using only anonymous communication is insufficient; an adversary can infer a user's identity from her or his partial information exposed in the network. This attack links the partial infor- mation, called a quasi-identifier [9], to the auxiliary information collected from other channels, including Web or public records. For example, Sweeney showed that 87% (216 million of 248 million) of the United States' population can be uniquely identified based only on ZIP code, gender, and date of birth, taken from publicly available data [26]. Golle recently updated this result. He showed that the quasi-identifier can uniquely identify 63% of the population of the United States [10].

For this problem, researchers have studied some approaches using public data releases. The main-stream approach is publishing anonymized datasets through randomized or cryptographic techniques [1, 23, 18] to add noise to the data records, thus achieving privacy goals such as k-anonymity so that a record cannot be distinguished from at least $k - 1$ other records. To effectively utilize this approach in social networks or personalized services where user data is quickly inserted, deleted, and modified, the following issues should be studied. First, the anonymization process performance cost may be huge, es- pecially for large and sparse databases. Second, maintaining anonymized datasets requires the expensive update cost, even if few records are newly inserted, deleted, or modified.

Because even simple personal data, including gender, race, medical condition or academic background, identified by linking auxiliary information can be bait that ultimately cheats users, the inference attack should be studied intensively as a main defense target.

6. CONCLUSION

Information technology in government services offers an interesting natural experiment. While commercial Web service providers compete for survival by increasing customers' confidence about security and privacy control technology, government Web services are typically provided unilaterally without allowing relevant user vulnerability evaluation.

The effects have been rather mixed. Sensitive contents from authoritative sources cause a cognitive bias that overestimates sensitivity and leads users to involuntarily authenticate sources. We demonstrated that exploiting this psychological response is an effective phishing strategy to cheat government service users. Our empirical analysis showed that this effect can be amplified using a private context in phishing messages, such as identification data discovered by inference from open online records.

The lessons learned from our experiment have much wider applications. We recommend that government Web service developers study the psychological effect that authoritative sources have on users. When constructing Web services, governmental characteristics should be presented differently than those of commercial service providers. Regarding links between a real identity and online personal data, online privacy control should be widely studied and adopted nationwide. This control requires close cooperation from users, the private sector and the government.

In future work, we intend to continuously conduct a threat and risk analysis on government Web ser- vices by studying attack trends and evaluating their severity and likelihood, especially from the linkability perspective. Few studies focus on the authoritative behaviors of governmental entities in administrative Web service development. We believe the threat model analysis helps determine the usable security re- quirements that enhance overall nationwide privacy protection levels and that will gain users' trust of Internet technologies. This analysis will be crucial for government Web services to succeed.

REFERENCES

[1] Rakesh Agrawal and Ramakrishnan Srikant. Privacy-preserving data mining. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (SIGMOD '00), pages 439–450, New York, NY, USA, 2000. ACM.

[2] Hege K Andreassen, Maria M Bujnowska-Fedak, Catherine E Chronaki, Roxana C Dumitru, Iveta Pudule, Silvina Santana, Henning Voss, and Rolf Wynn. European citizens' use of E-health services: a study of seven countries. BMC Public Health, 7(1):53–, 2007.

[3] Robert D. Atkinson and Daniel D. Castro. Digital quality of life. Available from: http://www.itif.org/files/DQOL.pdf, October 2008. [cited 2011 Apr 5].

[4] France Bélanger and Lemuria Carter. Trust and risk in e-government adoption. J Strategic Inf Syst, 17:165–176, 2008.

[5] Joseph Bonneau, Jonathan Anderson, and George Danezis. Prying data out of a social network. In Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, pages 249–254, Washington DC, USA, 2009. IEEE Computer Society.

[6] Joseph Bonneau and Sren Preibusch. The privacy jungle: on the market for data protection in social networks. In Tyler Moore, David Pym, and Christos Ioannidis, editors, Economics of Information Security and Privacy, pages 121–167. Springer US, 2010.

[7] Lemuria Carter and France Bélanger. The utilization of e-government services: citizen trust, inno- vation and acceptance factors. Inform Syst J, 15(1):5–25, 2005.

[8] Meghan E. Cook. What Citizens Want from E-Government. Available from: http://www.ctg.albany.edu/publications/reports/what_citizens_want/what_citizens_want.pdf, Octo- ber 2000. [cited 2011 Apr 5].

[9] Tore Dalenius. Finding a needle in a haystack or identifying anonymous census records. J Official Statistics, 2(3):329–336, 1986.

[10] Philippe Golle. Revisiting the uniqueness of simple demographics in the US population. In WPES '06: Proceedings of the 5th ACM Workshop on Privacy in Electronic Society, pages 77–80. ACM, 2006.

[11] Jeremiah Grossman. Cross-Site Scripting worms & viruses: the impending threat & the best defense. Available from: https://www.whitehatsec.com/home/assets/WP5CSS0607.pdf, June 2007. [cited 2011 Apr 27].

[12] Tom N. Jagatic, Nathaniel A. Johnson, Markus Jakobsson, and Filippo Menczer. Social phishing. Commun. ACM, 50(10):94–100, 2007.

[13] Markus Jakobsson. Modeling and preventing phishing attacks. In Andrew Patrick and Moti Yung, editors, Financial Cryptography and Data Security, volume 3570 of Lecture Notes in Computer Science, pages 578–578. Springer Berlin / Heidelberg, 2005.

[14] Hyoungshick Kim, Jun Ho Huh, and Ross Anderson. On the security of internet banking in south korea. Technical Report RR-10-01, Department of Computer Science, University of Oxford, March 2010.

[15] Tong Hyung Kim. Experts say specific tech mandates make internet banking vulnerable. Avail- able from: http://www.koreatimes.co.kr/www/news/biz/2010/05/123_65102.html, April 2010. [cited 2011 Apr 27].

[16] K Kita, JS Lee, H Suzuki, N Taira, M Yachida, H Yamamoto, Y Homma, T Obi, M Yamaguchi, and N Ohyama. The personal health information reference system based on e-P.O.Box conception. J Korean Soc Med Inform, 14(3):213–220, 2008.

[17] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. School of phish: a real-world evaluation of anti-phishing training. In Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS ’09), pages 3:1–3:12, New York, NY, USA, 2009. ACM.

[18] Kristen LeFevre, David J. DeWitt, and Raghu Ramakrishnan. Incognito: efficient full-domain K- anonymity. In Proceedings of the 2005 ACM SIGMOD international conference on Management of data (SIGMOD ’05), pages 49–60, New York, NY, USA, 2005. ACM.

[19] Margaret A. Winker, Annette Flanagin, Bonnie Chi-Lum, John White, Karen Andrews, Robert L. Kennett, Catherine D. DeAngelis, and Robert A. Musacchio. Guidelines for medical and health information sites on the Internet: principles governing AMA Web sites. JAMA-J. Am. Med. Assoc., 283(12):1600–1606, 2000.

[20] David L. McClure. Electronic government: challenges must be addressed with effective leadership and management. Available from: http://www.gao.gov/new.items/d01959t.pdf, July 2001. [cited 2011 Apr 5].

[21] Microsoft. Description of ActiveX technologies. Available from: http://support.microsoft.com/ kb/154544, January 2007. [cited 2011 Apr 27].

[22] Yunsang Oh, Takashi Obi, Joong Sun Lee, Hiroyuki Suzuki, and Nagaaki Ohyama. Empirical analysis of internet identity misuse: case study of south korean real name system. In Proceedings of the 6th ACM workshop on Digital identity management (DIM ’10), pages 27–34, New York, NY, USA, 2010. ACM.

[23] Benny Pinkas. Cryptographic techniques for privacy-preserving data mining. ACM SIGKDD Explo- rations Newsl, 4:12–19, December 2002.

[24] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The emperor’s new security indicators. In Proceedings of the 2007 IEEE Symposium on Security and Privacy (SP ’07), pages 51–65, Washington, DC, USA, 2007. IEEE Computer Society.

[25] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In Proceedings of the 3rd symposium on Usable privacy and security (SOUPS ’07), pages 88–99, New York, NY, USA, 2007. ACM.

[26] Latanya Sweeney. k-anonymity: a model for protecting privacy. Int J Uncertain Fuzz, 10(5):557–570, 2002.

BIBLIOGRAPHY OF AUTHORS

Yunsang Oh received the B.S. and M.S. degrees in Computer Science and Engineering from Sogang University, Korea in 1997 and 2002, respectively. During 2003-2009, he worked in Samsung Electronics. He also served a member of OMA for DRM standardization in mobile networks. He is currently studying at Tokyo Inst. of Tech, Japan as a Ph.D. student.

Takashi Obi received his M.E. and Ph.D. degree in information processing from Tokyo Inst. of Tech, Tokyo, Japan, in 1992 and 1997, respectively. He was a Research Associate (1995-2002) at Imaging Science and Engineering Lab., Tokyo Inst. of Tech. and a Visiting Researcher (1998, 2000) at Univ. of Pennsylvania. He has been an Associate Professor at Dept. of Information Processing, Tokyo Inst. of Tech. since 2002.