Phishing Susceptibility: The Good, the Bad, and the Ugly

Conference Paper · September 2016

DOI: 10.1109/ISI.2016.7745462

Ahmed Abbasi
University of Notre Dame
89 PUBLICATIONS **5,365** CITATIONS

SEE PROFILE

Yan Chen
Florida International University
40 PUBLICATIONS **1,117** CITATIONS

SEE PROFILE

Phishing Susceptibility: The Good, the Bad, and the Ugly

Ahmed Abbasi
IT Area and Center for Business Analytics
McIntire School of Commerce
University of Virginia
Charlottesville, VA, USA
abbasi@comm.virginia.edu

F. Mariam Zahedi
ITM Department
Sheldon B. Lubar School of Business
University of Wisconsin-Milwaukee
Milwaukee, WI, USA
zahedi@uwm.edu

Yan Chen
Department of Information Systems
College of Business
University of Auburn-Montgomery
Montgomery, AL, USA
ychen3@aum.edu

Abstract—**Phishing website-based attacks remain pervasive, with high user susceptibility continuing to be a major factor. In this study we use cluster analysis coupled with an elaborate controlled experiment involving hundreds of participants to identify and examine high susceptibility user segments in terms of their perceptions, demographics, and phishing website traversal behavior. The results reveal three sets of users, including two sets that exhibit highly problematic behavior. The results have important implications for training programs, usability of anti-phishing tools, and security policies.**

Keywords—phishing susceptibility; online security; Internet fraud; anti-phishing tools; security perceptions.

I. Introduction

Phishing is a type of semantic attack that exploits human vulnerabilities as opposed to vulnerabilities in hardware or software [2]. Over the past twenty years, there has been a fair amount of research on the design and development of better anti-phishing tools that employ rules, blacklists, and/or elaborate machine learning algorithms for superior phishing website detection performance [1][3][6][14]. Consequently, over that time period state-of-the-art phishing website detection rates have climbed from 60% to well above 90% [1][13][14]. However, user susceptibility to phishing attacks persists. Year in, year out, it remains one of the three most prevalent forms of cybercrime in both organizational and individual/household settings. In light of greater recognition that anti-phishing is a socio-technical problem where tool usability and user perception considerations must also be taken into account, several experiments and user studies have been conducted in recent years [4][5][10][16][17]. These studies have consistently come to the following two conclusions: (1) users have a hard time identifying phishing websites; (2) even when using anti-phishing tools, users routinely disregard tool warnings resulting in high levels of susceptibility. There remains a need for research to help uncover factors attributable to user susceptibility to phishing attacks. Such work is essential to improve security training curriculum [8], design more intuitive anti-phishing warnings [7][10], and identify high-risk user groups.

Recent work has called for greater use of predictive and descriptive analytics to understand the user-centric challenges pertaining to cyber-security [12][18]. Accordingly, in this study we employ cluster analysis in conjunction with an elaborate controlled experiment to examine how the interplay between user perceptions, demographics, and phishing website traversal behavior can offer insights regarding phishing susceptibility.

II. Related Work and Research Gaps

Two major categories of phishing websites are concocted and spoof sites. Spoof sites are replicas imitating existing websites. Concocted websites deceive users by attempting to appear as unique, legitimate entities. Anti-phishing tools are designed to protect users against phishing attacks that rely on spoof or concocted websites. Prior work has examined two important areas related to phishing: (1) the design, development, and evaluation of novel anti-phishing tools; (2) analysis of users' ability to detect and avoid phishing websites.

Existing anti-phishing tools use manually crafted rules, machine learning algorithms, and/or blacklists to detect phishing websites. Some even incorporate elaborate visual similarity or network stack variables [11][15]. These tools are often plugged into web browsers such as Internet Explorer, Chrome, Firefox, etc. or part of anti-virus software offered by companies such as McAfee, Symantec, and Kaspersky. The phishing detection rates for existing tools range from 60% for browser "look-up" tools that rely on online community-based blacklists, to over 90% for state-of-the-art machine learning based anti-phishing tools [1][3][6][9][13][14].

The bigger issue that has emerged is Internet users' inability to avoid phishing threats. Earlier studies found that as many as 70% of phishing websites were not identified by users [5][17]. Perhaps more alarming is susceptibility rates even when aided by an anti-phishing tool – with users disregarding tool warnings in 20% to 40% of instances [4][14][16], resulting in high susceptibility rates even when using accurate protective tools. Hence, there is a need to better understand what user factors are responsible, and which user segments are at the greatest risk of susceptibility to phishing attacks [19][21].

Accordingly, in this study the key questions we seek to answer using cluster analysis are: **(1)** *Which user segments are most susceptible to phishing website attacks?* **(2)** *What perceptual and demographic factors differentiate high-susceptibility segments from those less likely to fall for phishing attacks?*

III. Experiment design

A controlled lab experiment was use for data collection. In the experiment, participants were asked whether they would consider opening a savings account from a set of 10 online bank URLs that were provided. Online banking was chosen as the domain since financial institution websites are amongst the most commonly targeted by phishing attacks [14][22]. The URLs were displayed as search results, with 5 legitimate and 5 phishing websites presented to each participant. Both categories of phishing websites were incorporated: concocted and spoof [1][16]. Upon clicking on a URL, an anti-phishing tool warning appeared for certain phishing website instances. The tool had either 60% or 90% phishing detection rate. These two rates were incorporated since prior benchmarking studies have found that most anti-phishing tools' detection rates fall within this range [1][6][13]. Since false positive rates for most state-of-the-art industry tools are negligible (i.e., they rarely consider legitimate websites to be phish), warnings didn't appear for the legitimate websites [14]. The experiment was run using a between subject combinatorial design, where each participant was given either legitimate and spoof websites, or legitimate and concocted. Similarly, each participant was given an anti-phishing tool with either a 60% detection rate or a 90% detection rate. This resulted in 4 possible combinations. Further details regarding the experiment design are as follows.

Each participant was given 5 legitimate and 5 phishing URLs. The legitimate websites were roughly balanced between those with very high levels of web traffic, those with average levels, and those with below average traffic (based on Alexa traffic rank). Each participant was also given URLs for either 5 spoof or 5 concocted websites (always one of the two categories). These URLs were selected randomly from a set of 15 spoof and 15 concocted websites. The 15 spoof bank websites in the pool were replicas of the 15 legitimate websites, and were taken from the popular phishing database PhishTank. The concocted bank websites were taken from the Artists Against 4-1-9 database. The 10 URLs provided to each participant were displayed in random order [16].

The participants were required to click on each URL, in any order, to potentially browse the online banking website and examine their online savings account offerings. Clicking on a URL triggered the anti-phishing tool, which evaluated the website and made a recommendation. If the anti-phishing tool considered the website legitimate, the URL's page was displayed in the browser. On the other hand, if the tool deemed the site to be a phish, the participant's web browser was redirected to a warning page before they had an opportunity to browse the target website. The default Microsoft Internet Explorer warning page was used since it is similar to ones used by other popular browsers such as Mozilla Firefox and Google Chrome (these browsers collectively account for about 80% of the total desktop browser market share). When encountering a warning, participants had to decide whether to:

- Adhere to the warning and avoid continuing to the target website. In order to choose this option, participants had to click on a link on the warning page. This action would cause them to return to the URL list without having visited the potential phishing website.
- Disregard the warning and continue on to the website, anyway. In order to choose this option, the participants had to click on a hyperlink on the warning page. Consistent with the Microsoft Internet Explorer warning layout, the link to ignore the warning and proceed to the website was lower on the warning page and accompanied by a warning icon.

As previously discussed, in line with prior benchmarking studies [1][3][13], the anti-phishing tool's phishing detection rate was either 60% or 90%. In other words, the 90% tool was had a 10% likelihood of failing to display a warning for a phishing website URL. For each of the 30 phishing online banking websites in the experiment, the average misclassification rate across five popular anti-phishing tools was computed [1][13][14]. The likelihood of a given phishing URL being misclassified was proportional to this misclassification probability. This was done to ensure realism in regards to the anti-phishing tool misclassifications in our experiment – with phishing URLs easily detected by popular tools less likely to trigger an "alarm failure" in the experiment.

For each URL, after being exposed to the anti-phishing tool's recommendation, the task required participants to decide if they would visit the website, and if visited, to browse the website to evaluate savings account options. The participant's decisions regarding these two tasks were computed automatically using web analytics tracking software. Additionally, the participants needed to answer two questions: (1) whether they considered the website to be legitimate or a phishing site; (2) whether they would open a savings account (i.e., transact) with the website. Participants were evaluated based on their decisions to differentiate legitimate websites from phish [5][11][17], decisions to visit phishing websites, and willingness to transact with phishing sites [5]. Each participant had 20 minutes to make all their decisions regarding the 10 websites assigned to them. Pre-testing revealed the time allotted to be appropriate for the experiment tasks.

The experiment participants were 520 university students, staff, and members of the general public from a two cities in the United States. Each participant was randomly assigned to one of the four experiment settings: high-accuracy tool + spoof, high-accuracy tool + concocted, low-accuracy tool + concocted, and low-accuracy tool + spoof. Each of the four settings had roughly the same number of participants; around 127. Prior to the experiment, participants were given instructions regarding the aforementioned experiment task and were also made aware of the phishing detection rate associated with their particular anti-phishing tool. Prior to the experiment, each participant was also given a survey encompassing questions related to anti-phishing tool perceptions, threat perceptions, and their prior web experiences. Tool perception questions included whether they considered anti-phishing tools to be useful, and their perceptions regarding the cost of tool errors. The threat perception questions related to their awareness of phishing threats. Prior experience questions focused on their web reliance, trust and familiarity with online banks, familiarity with the specific bank URLs used in the experiment, and their past phishing encounters and losses.

IV. Cluster analysis Results

The results for 11 participants were omitted since they completed the experiment too quickly. The data from the remaining 509 participants, each of whom interacted with 5 phishing websites, was incorporated. This resulted in 2545 total user-phish interaction data points. The tool information and threat characteristic variables were excluded since the objective of the cluster analysis was to profile user-phish interactions based on users' experiences, demographics, perceptions, and phishing website traversal decisions. The clustering was performed using x-means, which utilizes the Bayesian information coefficient to determine the optimal number of clusters [23]. In order to triangulate the clustering results, k-means and expectation maximization were also used with the elbow method for determining the number of clusters. All three methods yielded comparable results in terms of number of clusters, cluster composition, and cluster centroid values for variables. Accordingly, we report the x-means clustering results here.

Table I depicts the cluster centroids. In total, six unique user-phish encounter clusters were generated. In the table, only variables with significant differences across clusters are shown. Consistent with prior work, variables were normalized to a 0-1 scale prior to clustering to avoid excessively weighting variables with larger value ranges. Accordingly, all centroid values also range between 0 and 1. Significances in differences between variable means across clusters were calculated using pair-wise t-tests (with all p-values < 0.05 considered significant). Figure 1 depicts the phishing website traversal decisions pertaining to these six clusters, include the percentage of cluster instances that visited, browsed, considered legitimate, and intended to transact with phishing websites, respectively.

Based on the results appearing in Table 1 and Figure 1, from a phishing susceptibility perspective, the six clusters can be grouped into three categories: the good, the bad, and the ugly. The "good" clusters were Clusters 1 and 2. Only 28.9% of instances in Cluster 1 visited a phishing website, with only 3.3% of user-phish encounters ending with intention to transact with a phishing website. Looking at significant values in the centroid table, members of this cluster were more likely to be older, educated, female, with higher perceptions regarding the usefulness of anti-phishing tools, and greater familiarity with online banks. Cluster 2 has similarly low phishing website traversal rates, with 35.2% of phishing URL encounters resulting in a visit, but only 3.2% being considered legitimate and 2.6% ending with user intention to transact with a phishing website. The two key characteristics for this cluster were that they had the highest perceived usefulness for the anti-phishing tool, and the highest past losses attributable to phishing.

The "bad" clusters were Clusters 3 and 4. As depicted in Figure 1, relative to the "good" clusters, members of Clusters 3 and 4 were far more likely to visit and browse phishing websites (with visitation/browsing rates between 89.5% and 98.8%). Their intention to transact rates (between 6.3% and 16.1%), while lower relative to their visit and browse rates, were still two to four times higher than those pertaining to the "good" clusters. In terms of clusters centroids, there was a clear dichotomy between Clusters 3 and 4. Cluster 3 encompassed those that were younger, with high phishing awareness, and more prior encounters with phishing attacks. Conversely, Cluster 4 was comprised of those with lower phishing awareness, past encounters and losses, but greater web reliance.

As depicted in Figure 1, the "ugly" Clusters (5 and 6) were ones that not only had high visit and browse rates, but were also very likely to consider phishing websites as legitimate (91.1% of cluster 5 and 68.1% of cluster 6). Not surprisingly, they also had the highest intention to transact rates. In particular, 66% of instances in Cluster 6 expressed an intention to transact with the phishing website they encountered. The most common theme with respect to these two clusters was their low perceptions of anti-phishing tool usefulness. This resulted in high disregard rates which contributed to the extensive phishing website traversal percentages and risky decisions.

TABLE I. Centroids for six susceptibility clusters

Category	Variable	Good		Bad		Ugly	
		1	2	3	4	5	6
Tool Perceptions	Tool Usefulness	0.590^{*}	0.594^{*}	0.515	0.538	0.491^{+}	0.492^{+}
	Cost of Tool Error	0.437	0.385	0.466	0.502	0.497	0.527^{**}
Threat Perceptions	Phishing Awareness	0.473	0.512	0.552^{**}	0.439^{++}	0.499	0.491
Demographics	Gender	1.000	0.000	0.000	1.000	0.416	0.378
	Age	0.154^{**}	0.101	0.087^{++}	0.112	0.099	0.103
	Education	0.413^{**}	0.375	0.368	0.382	0.374	0.378
Prior Web Experiences	Web Reliance	0.722	0.707	0.704	0.765^{**}	0.710	0.724
	Trust in Institution	0.534	0.509	0.531	0.532	0.526	0.557^{**}
	Familiarity w/ Domain	0.345^{**}	0.304	0.263	0.288	0.278	0.280
	Familiarity w/ Site	0.082	0.067	0.090	0.074	0.058^{++}	0.164^{**}
	Past Losses	0.142	0.161^{**}	0.130	0.100^{++}	0.130	0.144
	Past Encounters	0.164	0.194	0.205^{**}	0.128^{++}	0.176	0.185
Number of User-Phish Instances Per Cluster		322	384	664	410	271	494

** Significantly higher than all other clusters; * Significantly higher than all 4 clusters within the two other groups
++ Significantly lower than all other clusters; + Significantly lower than all 4 clusters within the two other groups

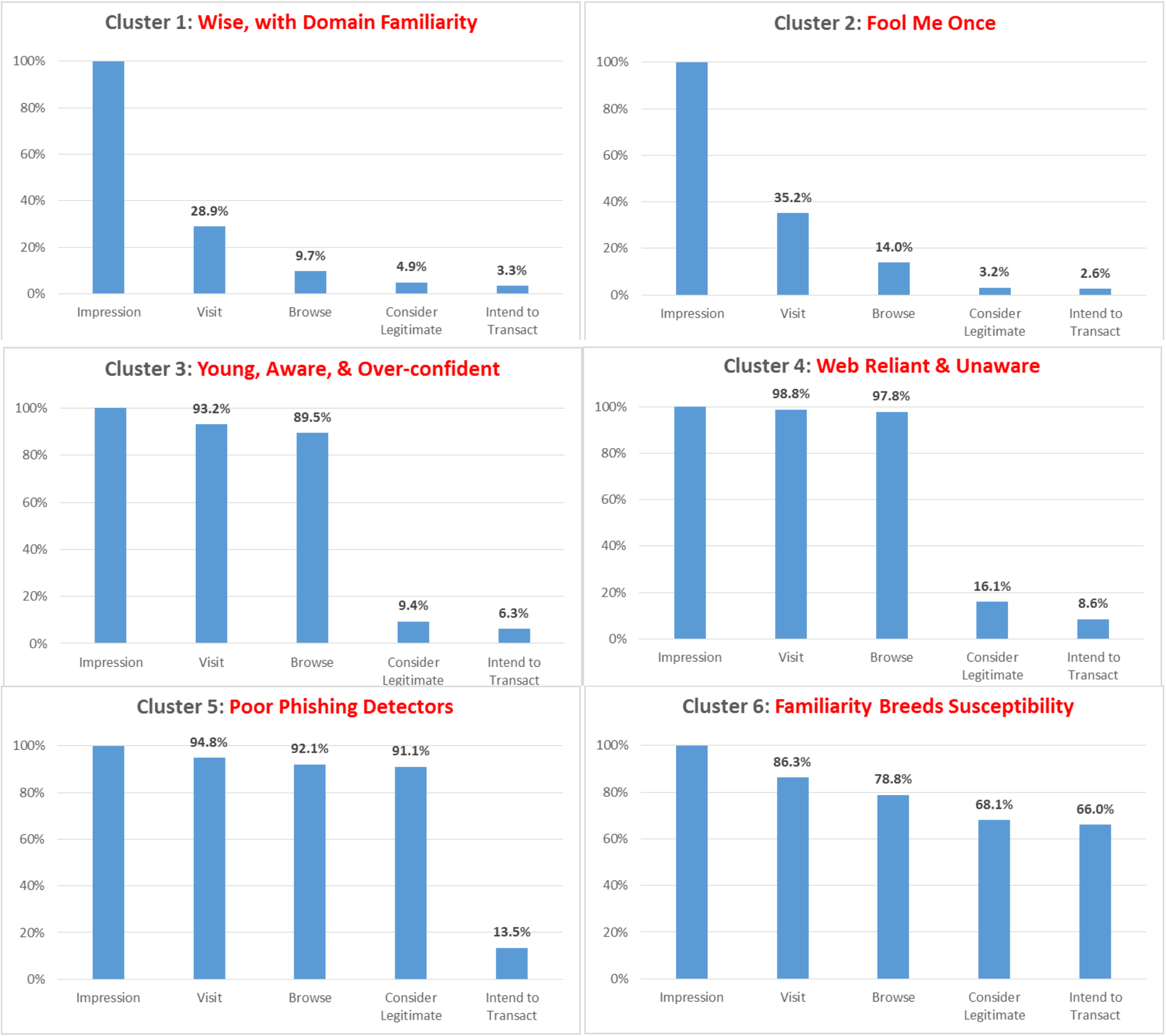

Fig. 1. Phishing Website Traversal Statistics for Six User-Phish Instance Clusters

Table II depicts cluster personas and descriptions developed based on the cluster centroid and phishing website traversal statistics presented in Table I and Figure I. Regarding the "good" clusters, Cluster 1 included females with an average age of 28, high familiarity with online banks, high perceived tool usefulness, and risk averse behavior with respect to traversal of phishing websites and legitimacy/transacting decisions. Accordingly, we labeled this cluster, which accounted for 12.7% of all user-phish instances in our data, as "Wise, with Domain Familiarity." Cluster 2 encompassed males that had suffered significantly higher past losses, and consequently had lower trust in online banks. This cluster, which accounted for 15% of the participants, was the least likely to consider phishing websites legitimate, or to transact with such websites. We labeled this cluster "Fool Me Once."

The "bad" clusters contained users with high visit and browse likelihoods. Cluster 3, labeled "Young, Aware, & Over-confident," consisted of younger males with significantly higher past encounters and higher perceived phishing awareness. Though not included in the cluster analysis, they also had significantly higher self-efficacy. Consequently, this group was far more likely to visit and browse phishing websites. Unfortunately, on average, they were also at least twice as likely to consider such sites legitimate (and to transact with them), relative to members of the "good" clusters. Cluster 4, "Web Reliant and Unaware," were females with high web reliance, little past losses, and low phishing awareness. The lack of prior experience and encounters resulted in a markedly wider funnel than members of the "good" clusters. The "ugly" clusters comprised of participants with heavy traversal of phishing websites and risky behavior. As alluded to, both Clusters 5 and 6 had significantly lower perceived tool usefulness. As a result, these groups routinely disregarded tool warnings, resulting in 91% and 68% of phishing websites considered legitimate, respectively. Accordingly, Cluster 5 was labeled "Poor Phishing Detectors." In the case of Cluster 6, "Familiarity Breeds Susceptibility," significantly higher familiarity with the sites coupled with trust in banks resulted in transaction intention rates of 66%.

TABLE II. SUMMARY OF KEY USER PHISHING SUSCEPTIBILITY CLUSTERS

Cluster	Description	Key Characteristics
	The Good	
1. Wise, with Domain Familiarity (12.7%)	This cluster was comprised of females that had an average age of 28, with 25% over the age of 31. In addition to being the most educated, this group had the highest familiarity with the domain (online banks) and high perceived tool usefulness. With respect to phishing website traversal behavior, they were the least likely to visit and browse phishing websites. Less than 5% of phish were considered legitimate, with 3.3% resulting in intention-to-transact.	Traversal: ▼All stages Personal Attributes: ▲Age ▲Female ▲Familiarity with Domain ▲Tool Usefulness
2. Fool Me Once (15.1%)	This group consisted of males that had suffered higher past losses than other groups. Consequently, they were less trusting of bank/pharmacy websites. The group also had the highest perceived tool usefulness and the lowest perceived cost of tool error. This group's reliance on the tool resulted in the lowest consider-legit and intend-to-transact rates.	Traversal: ▼All stages Personal Attributes: ▲Male ▲Past Losses ▲Tool Usefulness ▼Trust in Institution
	The Bad	
3. Young, Aware, & Over-confident (26.1%)	This group consisted of younger males, with an average age of 22 (45% at or below 20 years of age). The group had the greatest past encounters with phishing websites (but with slightly below average losses), and consequently, the highest reported phishing awareness. Though not included in the cluster analysis, they also had significantly higher self-efficacy. This confidence resulted in high visit and browse rates, and consider-legit and intention-to-transact rates of 9.4% and 6.3%, respectively.	Traversal: ▲Visit, Browse ◄Consider-legit, Transact Personal Attributes: ▲Male ▼Age ▲Past Encounters ▲Phishing Awareness
4. Web Reliant & Unaware (16.1%)	This group consisted of females with the highest reported reliance on the Web. The group also had the lowest past losses, past encounters, and overall phishing awareness. Consequently, the group had high visitation rates, the highest browse rates, and consider-legit and intention-to-transact rates of 16.1% and 8.6%, respectively.	Traversal: ▲Visit, Browse ◄Consider-legit, Transact Personal Attributes: ▲Female ▲Web Reliance ▼Past Losses ▼Phishing Awareness
	The Ugly	
5. Poor Phishing Detectors (10.6%)	This group had the lowest perceived tool usefulness, choosing to rely almost entirely on intuition for funnel-related decisions. Unfortunately, this resulted in high visit and browse rates, and a whopping 91% of phishing websites being deemed as legitimate. Furthermore, 13.5% of interactions ended with an intention-to-transact.	Traversal: ▲First three stages ◄Transact Personal Attributes: ▼Tool Usefulness ▼Familiarity with Site
6. Familiarity Breeds Susceptibility (19.4%)	This group had the highest familiarity with the phishing sites, and the highest trust in banking websites. Much like the "Poor Phishing Detectors" group, they also had low perceived tool usefulness and the highest perceived cost of tool error. The combination of familiarity/trust and lack of tool usage resulted in a consider-legit rate of 68% and 66% of interactions with intention-to-transact.	Traversal: ▲All stages Personal Attributes: ▲Familiarity with Site ▲Trust in Institution ▼Tool Usefulness ▲Cost of Tool Error

V. Results Discussion and Conclusions

Overall, the cluster analysis underscores the importance of segmenting users based on their phishing susceptibility characteristics, including key perceptual and demographic factors as well as phishing website traversal and risky behavior tendencies. The clusters shed light on how phishers are able to exploit vulnerabilities attributable to users' phishing awareness, anti-phishing tool perceptions, familiarity with certain sites, web reliance, and demographic characteristics. The results have important implications for several stakeholder groups. In the context of training and education programs, clearly improving user perceptions regarding perceived usefulness of the tools must be a focal point, considering that it was a major differentiator between the "good" and "ugly" clusters. Furthermore, adding modules designed to prevent users' institutional trust and site familiarity from being used against them is an important direction [4]. Increasing user phishing awareness, particularly amongst those that are highly web reliant and lack exposure to phishing, could be beneficial (e.g., Cluster 4). Conversely, those with awareness and encounters (e.g., Cluster 3) should be provided with more advanced examples to help reduce the "over-confidence" factor. The results could also be used for custom access control policies and/or anti-phishing tool settings, particularly for high-risk user clusters. Some of the insights might also be applicable to customized anti-phishing warning design, where users' cluster affiliations and personal tendencies could be used to determine the best manner in which to warn that profile of users [7][20].

In this study, we used cluster analysis coupled with an elaborate controlled user experiment to examine users' phishing susceptibility tendencies. The cluster analysis revealed six clusters associated with three key segments. We discussed each segment's significant tendencies, their phishing traversal behavior, and resulting personifications. The results clearly identify two sets of problematic user segments: those that visit, browse, and consider legit, and those with are also highly likely to consider phishing websites legitimate. Moreover, these two sets (i.e., the "bad" and "ugly" clusters) account for nearly two-thirds of users in our experiments. These finding have important implications for security programs, training, and custom access control policies. Given that phishing is a semantic attack that exploits human vulnerabilities, future anti-phishing research should continue to adopt a socio-technical perspective such that Internet users, the last line of defense, are better equipped to avoid phishing attacks.

Acknowledgment

This research has been supported in part by the following two U.S. National Science Foundation grants: CNS-1049497 "A User-Centric Approach to the Design of Intelligent Fake Website Detection Systems," and ACI-1443019 "DIBBs for Intelligence and Security Informatics Research Community."

References

[1] A. Abbasi and H. Chen, "A Comparison of Tools for Detecting Fake Websites," IEEE Computer, vol. 42, pp. 78-86, October 2009.

[2] B. Schneier, "Semantic Network Attacks," Comm. of the ACM, vol. 43, p. 168, December 2000.

[3] A. Abbasi and H. Chen, "A Comparison of Fraud Cues and Classification Methods for Fake Escrow Website Detection," Information Technology and Management, vol. 10(2), pp. 83-101, 2009.

[4] M. Wu, R. C. Miller, and S. L. Garfunkel, "Do Security Toolbars Actually Prevent Phishing Attacks?," In Proc. Conf. on Human Factors in Computing Systems, pp. 601-610, 2006.

[5] S. Grazioli and S. L. Jarvenpaa, "Perils of Internet Fraud: An Empirical Investigation of Deception and Trust with Experienced Internet Consumers," IEEE Trans. Systems, Man, and Cybernetics Part A, vol. 20(4), pp. 395-410, 2000.

[6] A. Abbasi, Z. Zhang, D. Zimbra, H. Chen, and J. F. Nunamaker Jr., "Detecting Fake Websites: The Contribution of Statistical Learning Theory," MIS Quarterly, vol. 34(3), pp. 435-461, 2010.

[7] Y. Chen, F. M. Zahedi, and A. Abbasi, "Interface Design Elements for Anti-phishing Systems," In Proc. Intl. Conf. Design Science Research in Information Systems and Technology, pp. 253- 265, 2011.

[8] P. Kumaraguru, S. Sheng, A. Aquisti, L. F. Cranor, and J. Hong, "Teaching Johnny Not to Fall for Phish," ACM Trans. Internet Technology, vol. 10(2), no. 7, 2010.

[9] A. Abbasi, F. M. Zahedi, and S. Kaza "Detecting Fake Medical Websites using Recursive Trust Labeling," ACM Trans. Information Systems, 30(4), article 22, 2012.

[10] L. Li and M. Helenius, "Usability Evaluation of Anti-Phishing Toolbars," J. Computer Virology, vol. 3(2), pp. 163-184, 2007.

[11] W. Liu, X. Deng, G. Huang, and A. Y. Fu, "An Antiphishing Strategy Based on Visual Similarity Assessment," IEEE Internet Computing, vol. 10, pp. 58-65, February 2006.

[12] A. Abbasi, W. Li, V. A. Benjamin, S. Hu, and H. Chen, "Descriptive Analytics: Examining Expert Hackers in Web Forums," In Proc. Joint IEEE Intelligence and Security Informatics Conf., pp. 56-63, 2014.

[13] Y. Zhang, S. Egelman, L. F. Cranor, and J. Hong, "Phinding Phish: Evaluating Anti-phishing Tools," In Proc. 14th Annual Network and Distributed System Security Symposium (NDSS), 2007.

[14] A. Abbasi, F. M. Zahedi, D. Zeng, Y. Chen, H. Chen, and J. F. Nunamaker Jr, "Enhancing Predictive Analytics for Anti-phishing by Exploiting Website Genre Information," Journal of Management Information Systems, 31(4), pp. 109-157, 2015.

[15] J. Koepke, S. Kaza, A. Abbasi, "Exploratory Experiments to Identify Fake Websites by using Features from the Network Stack," In Proc. IEEE Intl. Conf. on Intelligence and Security Informatics, pp. 126-128, 2012.

[16] A. Abbasi, F. M. Zahedi, and Y. Chen, "Impact of anti-phishing tool performance on attack success rates," In Proc. IEEE Intl. Conf. on Intelligence and Security Informatics, pp. 12-17, 2012.

[17] R. Dhamija, J. D. Tygar, and M. Hearst, "Why Phishing Works," In Proc. ACM Intl. Conf. Computer Human Interaction, pp. 581-590, 2006.

[18] A. Abbasi, S. Sarker, R. H. L. Chiang, "Big Data Research in Information Systems: Toward an Inclusive Research Agenda," Journal of the Association for Information Systems, 17(2), article 3, 2016

[19] F. M. Zahedi, A. Abbasi, and Y. Chen, "Design Elements that Promote the use of Fake Website Detection Tools," In Proc. 10th Annual AIS SIG-HCI Workshop, 2011.

[20] A. Herzberg and A. Jbara, "Security and identification indicators for browsers against spoofing and phishing attacks," ACM Transactions on Internet Technology, 8(4), article 16, 2008.

[21] F. M. Zahedi, A. Abbasi, Y. Chen, "Fake-Website Detection Tools: Identifying Elements that Promote Individuals' Use and Enhance Their Performance," Journal of the Association for Information Systems, 16(6), pp. 448, 2015.

[22] D. Dobolyi and A. Abbasi "PhishMonger: A Free and Open Source Public Archive of Real-World Phishing Websites," Submitted to Proc. IEEE Intl. Conference on Intelligence and Security Informatics, 2016.

[23] D. Pelleg and A. Moore "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," In Proc. Intl. Conf. Machine Learning, pp. 727-734, 2000.