The language effect in phishing susceptibility

Joakim Kävrestad[a], Rickard Pettersson[a] and Marcus Nohlberg[a]

[a] University of Skövde, Högskolevägen 1, 541 28 Skövde, Sweden

Abstract

Phishing has been, and remains to be, one of the most common types of social engineering. It is the act of tricking users to perform actions they normally wouldn't't using e-mail. Since phishing involves using technical measures to trick users, it is a social technical phenomenon that must be understood from the technical as well as the social side. While phishing and phishing susceptibility has been researched for decades, the effect of language ability on phishing susceptibility is underresearched. In this paper, we conducted a survey where we had swedes rate their English ability before classifying e-mails in Swedish and English as fraudulent or legitimate. The results shows that the respondents English ability does affect the ability to correctly identify legitimate emails and brings another piece to the puzzle of phishing susceptibility.

Keywords

phishing, susceptibility, foreign, language

1. Introduction

Social engineering, the act of deceiving end-users, has become one of the most devastating attacks against computer systems. Attackers manipulate human users in order to circumvent technical security measures in the endeavour to get access to login credentials, social security numbers, credit card information or the system itself [1]. Humans are seen by attackers as the easiest way into a network and the human factor is involved in 95% of security incidents in companies [2].

Phishing is a common online threat that is one type of social engineering, and it has been around since 1995 [3]. Phishing is indeed a practice that has been used by attackers for a long time and organisations are activity trying to combat it using detection tools, information campaigns and user training. Even so, phishing attacks continue to be used by attackers that manage to be successful and causing millions of dollars in damages.

Phishing is a complex matter where technology is used to deceive users into performing actions they would not normally do, making is a Socio-technical system [4]. The phenomenon that does not appear to go away. User susceptibility to phishing is widely researched, and suggests that users are bad at recognizing phishing to a satisfactory degree [2, 5]. The aim of this study is to research a phishing susceptibility aspect that has not gotten a lot of previous attention from the research community; *how good Swedish users are at detecting phishing e-mails in their native language compared to in English.* As such, the paper responds to a need for greater understanding of the social elements of phishing as described necessary in the literature [6, 7, 8]. This factor is increasing in importance due to the multilingual nature of many organizations. For instance, many Swedish organizations now have English as the first language within the organization and the same is seen in many other nations.

The rest of this paper is structured as follows; Section 2 presents the methodology used in the study. Section 3 presents the results and Section 4 concludes the paper and provides directions for future

6th International Workshop on Socio-Technical Perspective in IS development (STPIS'20), June 08–09, 2020, Online

EMAIL: joakim.kavrestad@his.se (J. Kävrestad); marcus.nohlberg@his.se (M. Nohlberg)
ORCID: 0000-0003-2084-9119 (J. Kävrestad); 0000-0001-5962-9995 (M. Nohlberg)

CEUR Workshop Proceedings (CEUR-WS.org)

work.

2. Methodology

The study was carried out using an online survey, distributed using SurveyMonkey, in which Swedish social network users were asked to classify 32 e-mails as phishing or legitimate. The e-mails were grouped into four different groups, with eight e-mails in each, as follows:

- Legitimate Swedish Emails (LS)
- Fraudulent Swedish e-mails (FS)
- Legitimate English E-mails (LE)
- Fraudulent English E-mails (FE)

The participants were also asked to rate their English proficiency on a six-graded scale based on the CEFRL framework[9]. CEFRL is a guideline for describing achievements in a foreign language and divides language capability into six levels. The levels were presented to the survey respondents as follows (Translated from Swedish):

- Beginner - You can present yourself and use simple words and phrases
- Basic - You can understand phrases and the most common words and you can communicate in simple contexts
- Intermediate - You can handle most situations that arise during travels to countries where English is used
- Upper intermediate - You can understand the main parts of complex text and , to a certain degree, interact fluently and spontaneously
- Advanced - You can read and understand a large portion of long and demanding texts and use English spontaneously, flexibly and efficiently in social contexts
- You can, without problem, understand everything you read or hear and express yourself fluently in almost any situation

The survey was designed to mimic an authentic situation as far as possible and was therefore constructed with 32 e-mails that the respondents were asked to classify as phishing or legitimate. The participants were told that they would receive a score based on how many correct classifications they made to introduce an element of gamification, inspired by [10]. The e-mails used in the survey were collected from open sources on the Internet and designed to be hard to classify.The survey was subjected to pilot testing during one week to ensure that it was understandable to the participants. Once completed, the survey was spread on social networks.

For data analysis, the participants were separated based on their reported English ability. Respondents classifying themselves in one of the two highest grades were placed in one group (A), and the other respondents in another group (B). The separation was done arbitrarily to get as equal group sizes as possible. Mean and Median values for the the two groups and four variables were then calculated to describe central tendencies. As suggested by [11], Shapiro-Wilks tests were used to assess the distribution form in combination with visual inspection of the gathered data. The variables were

Category	Group	Mean	Median	Mann-Whitney	T-test
LS	A	5.32	5.5	0.08	0.31
LS	B	5.64	6		
FS	A	6.68	7	0.67	0.32
FS	B	6.54	7		
LE	A	4.73	5	0.000	0.000
LE	B	3.49	3		
FE	A	6.37	7	0.75	0.65
FE	B	6.47	7		

Table 1
Statistics overview

found to not be normally distributed and as such, the MannWhitney U-test was primarily used to assess if identified differences between groups was statistically significant [12]. To validated the result of the significance test, T-test was also used for the same purpose providing increased validity through triangulation [13]. T-test is parametric and not considered appropriate for data not normally distributed, but can be argued to be robust in this case given the sample size[14]. The conventional significance level of 95% (p<0.05) was used throughout this study.

To further analyze the correlation between the ability to detect phishing e-mails and perceived language skill, correlation testing was used for variables where statistically significant differences were identified. Because of the above-mentioned concerns with the distribution form, Kendall's Tau was used as the primary correlation test and the parametric Pearsons r was used for validation. Those tests return a value between -1 and 1 where 1 signifies a perfect positive correlation and -1 signifies a perfect negative correlation [15].

3. Results

The survey was answered by 152 respondents and the collected data was used to calculate 2 scores for each participants. The score reflected the number of correct classifications the participant made in the following categories, and was calculated as a number between 0 and 8:

- Legitimate Swedish Emails (LS)
- Fraudulent Swedish e-mails (FS)
- Legitimate English E-mails (LE)
- Fraudulent English E-mails (FE)

The respondents were grouped based on their perceived English ability into one of two groups. The respondents ranking themselves in one of the two highest categories were put in one group (A, n=78) and the participants ranking themselves in one of the four lowest categories were placed in group B (n=74). An overview of the mean and median scores in the different categories for the two groups is presented in Table 1, below.

The data shown in Table 1 shows that the participants perform well in the two fraudulent categories, with mean values around 6.5 (of 8). This shows that 81% of the e-mails that were phishing e-mails was accurately identified to be phishing in this study. the mean values indicate a very small difference (0.1) between the language groups and the p-values are far higher than 0.05 showing that the identified difference can very well be due to chance. As such, the study does not suggest that the perceived English ability impact the ability to correctly identify phishing e-mails.

As seen in table 1, participants with a perceived high English proficiency score higher when it comes to accurately identify legitimate e-mails (mean difference of 1.24) and slightly better for accurately identifying English phishing e-mail. Mann-Whitney U-test was used to determine if the observed tendencies were significant. A p-value of below 0.05 shows significance meaning that the result for legitimate English e-mails is statistically significant, the result is validated by the T-test.

For the variables were significant results was observed, correlation testing was performed. the correlation test allows for use of the full language proficiency scale and will account for nuances that can be missed using the arbitrarily assigned binary variable for language proficiency. The results of the correlations tests between the variables for English proficiency and ability to accurately classify legitimate English e-mails were as follows:

- Kendalls Tau: 0.228 (p=0.00)
- Pearsons r: 0.279 (p=0.00)

The tests produce positive numbers around 0.25 with a p-value of 0 meaning that a positive correlation is identified and is significant at the level adopted in this study. This suggests that English ability is correlated with ability of correctly identifying legitimate e-mails in English even if the correlation coefficient also suggests that other factors, beyond the scope of this study, plays a role.

4. Conclusions

The aim of this study was to identify how the perceived English ability of Swedish participants affect the participants ability to correctly identify e-mails as legitimate of phishing. A survey containing 32 e-mails in four different categories were used and participants were invited to classify the e-mails as fraudulent or legitimate. The participants got a score in each category, and were then grouped by their self reported English ability.

In summary, the survey suggests that Swedish speaking users that are good at English are better at correctly identifying legitimate English e-mail as legitimate. However, perceived language skill does not seem to impact the ability to detect phishing e-mails in this survey. Still, the survey does suggest that language proficiency is an important factor in determining if e-mails are legitimate or not, which should be considered in future attempts to prevent phishing.

Another insight from this survey is that the participants score rather high in terms of correctly identifying phishing e-mails, the mean score is around 6.5 for Swedish and English phishing e-mails showing that the mean percentage of correct answers was 81%. this is almost 10% higher than the results reported by [16] and far better than [17] and [2] where about 70% of the participants fell for the phishing experiments conducted. The difference in results can have various reasons. One can, of course, be that the sample examined in this study is better at phishing detection than samples in other studies. It is known that cultural aspects affect security behaviour[18, 19]. another explanation to the difference can be that this study, as well as [16] use a survey methodology where participants are aware that they are being tested on phishing and focused on finding phishing e-mails. [17] and [2] perform observation studies where the users normal use of computers is taken into account. This suggests that awareness is a major factor in a users ability to detect phishing e-mails. While this discussion questions the use of the survey methodology for phishing susceptibility studies, it is hard to see another suitable method for measuring differences between controlled groups. And while the success rate reported in this paper should perhaps be interpreted with care, the paper successfully identifies that language ability affects the ability to correctly identify legitimate e-mails as legitimate.

Another conclusion from this study is that English ability impacts the participants ability to correctly identify legitimate English e-mails but does not impacts the ability to identify phishing e-mails. This result could suggest that the participants use non-language related cues (e.g. senders address and link addresses) to identify phishing e-mails. As such, the study identifies an interesting conundrum. Looking for language mistakes is a common advice on how to detect fraudulent e-mails, but if you receive e-mail in a language that you are not fluent in it is perhaps not a very helpful advice. In light of the world and organizations in it becomes more global, one can argue for a need of better ways to assist users in detecting fraudulent e-mails.

This study contributes to the knowledge around phishing susceptibility and shows that language skill is an important factor when users identify e-mails as legitimate or fraudulent. The paper, in comparison to other phishing susceptibility studies, also suggest that awareness is a key factor in phishing susceptibility, an insight that contributes to practitioners that design and implement security measures. While phishing and susceptibility to phishing is well researched the problem remains. As discussed in this paper, phishing research is difficult not only due to the complex nature of the problem but also due to ethical restrictions. The need for future work focusing on identifying ways to combat phishing is imperative. Future studies could focus on finding ethically sound methods to perform in-dept studies on phishing susceptibility. Another direction for future work could be survey-based studies used to identify other demographic aspects that affect phishing susceptibility.

References

[1] F. Salahdine, N. Kaabouch, Social engineering attacks: A survey, Future Internet 11 (2019) 89.

[2] A. Diaz, A. T. Sherman, A. Joshi, Phishing in an academic community: A study of user susceptibility and behavior, Cryptologia 44 (2020) 53–67.

[3] K. L. Chiew, K. S. C. Yong, C. L. Tan, A survey of phishing attacks: their types, vectors and technical approaches, Expert Systems with Applications 106 (2018) 1–20.

[4] D. Lacey, P. Salmon, P. Glancy, Taking the bait: a systems analysis of phishing attacks, Procedia Manufacturing 3 (2015) 1109–1116.

[5] K. Parsons, A. McCormac, M. Pattinson, M. Butavicius, C. Jerram, Phishing for the truth: A scenario-based experiment of users' behavioural response to emails, in: IFIP International Information Security Conference, Springer, 2013, pp. 366–378.

[6] A. Ferreira, P. M. V. Marques, Phishing through time: A ten year story based on abstracts., in: ICISSP, 2018, pp. 225–232.

[7] E. J. Williams, J. Hinds, A. N. Joinson, Exploring susceptibility to phishing in the workplace, International Journal of Human-Computer Studies 120 (2018) 1–13.

[8] A. Abbasi, F. M. Zahedi, Y. Chen, Phishing susceptibility: The good, the bad, and the ugly, in: 2016 IEEE Conference on Intelligence and Security Informatics (ISI), IEEE, 2016, pp. 169–174.

[9] Council of Europe, Common European Framework of Reference for Languages: learning, teaching, assessment, 2018.

[10] M. L. Hale, R. F. Gamble, P. Gamble, Cyberphishing: A game-based platform for phishing awareness testing, in: 2015 48th Hawaii International Conference on System Sciences, IEEE, 2015, pp. 5260–5269.

[11] M. Mendes, A. Pala, Type i error rate and power of three normality tests, Pakistan Journal of Information and Technology 2 (2003) 135–139.

[12] P. E. McKnight, J. Najab, Mann-whitney u test, The Corsini encyclopedia of psychology (2010) 1–1.

[13] Y. S. Lincoln, E. G. Guba, Naturalistic inquiry, 1985.

[14] G. Norman, Likert scales, levels of measurement and the "laws" of statistics, Advances in health sciences education 15 (2010) 625–632.

[15] M. M. Mukaka, A guide to appropriate use of correlation coefficient in medical research, Malawi medical journal 24 (2012) 69–71.

[16] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, J. Downs, Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2010, pp. 373–382.

[17] A. Mihelič, M. Jevšček, S. Vrhovec, I. Bernik, Testing the human backdoor: Organizational response to a phishing campaign, Journal of Universal Computer Science 25 (2019) 1458–1477.

[18] K.-L. Thomson, R. Von Solms, L. Louw, Cultivating an organizational information security culture, Computer fraud & security 2006 (2006) 7–11.

[19] A. Da Veiga, J. H. Eloff, A framework and assessment instrument for information security culture, Computers & Security 29 (2010) 196–207.