2021

Applying Usability Methods to Categorization of Phishing Emails

Oshin T. Thomas
University of Central Florida

Recommended Citation

Thomas, Oshin T., "Applying Usability Methods to Categorization of Phishing Emails" (2021). *Honors Undergraduate Theses*. 941.
https://stars.library.ucf.edu/honorstheses/941

APPLYING USABILITY METHODS TO CATEGORIZATION OF PHISHING EMAILS

by

OSHIN TEENA THOMAS

A thesis submitted in partial fulfillment of the requirements
for the Honors Undergraduate Thesis in Psychology
in the College of the Sciences
and in the Burnett Honors College
at the University of Central Florida
Orlando, Florida

Spring Term
2021

Thesis Chair: Dr. Corey Bohil (Ph.D.)

ABSTRACT

Phishing emails are a continuing threat in today's society—this study aimed to unpeel the layers on why certain people are prone to phishing emails than others. Participants were presented with twenty legitimate emails as well as twenty phishing emails in random order and were asked to choose whether they were phishing or not. They were also asked to explain why they chose the answer they believed was right in a couple of sentences. Data was compiled and collected via a Qualtrics survey and analyzed using JASP. Results obtained indicated little to no correlation between the number of features mentioned in the study and classifying and email accuracy as phishing. Studies like this help understand the cognitive constructs that lead to people unknowingly falling in traps set out by phishers and what prompts susceptible victims to think such emails are legitimate rather than seeing the dangers behind them.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Living in this highly technological era, we are faced with an increase in the number of cybercrime attacks daily. One such cybercrime is online phishing, a particularly dangerous means of obtaining confidential information. It is defined as "a form of deception in which an attacker attempts to fraudulently acquire sensitive information from a victim by impersonating a trustworthy entity" (Jagatic, Johnson, Jakobsson, & Menczer, 2007). Loss of funds and identity thefts are common ways these fraudulent emails harm their victims. Phishing emails are commonly presented in ways that seem realistic. The sender's identity is almost always spoofed or deliberately faked so as to appear to be sent by a trusted individual or organization. Once an attacker has access to a system, they can often steal personally identifiable information and sell it for profit, encrypt a database and demand ransomware to unlock it, manipulate and falsify clinical data, disrupt system availability, or perform other malicious activities (Gordon, Fairhall,& Landman, 2017). Due to the fear of being victims of fraud, people lose their trust in Internet transactions and decrease Internet businesses (Leyden, 2004). For example, many people believe using online banking increases the likelihood that they will become victims of identity theft, even though online banking provides more secure identity protection than paper and mail-based systems (Van Dyke, 2004).

Phishing has become a very grave global security issue. Verizon's 2017 Data Breach Investigations reported that almost all successful phishing attacks in 2016 were after malware installation, with 66% of all malware installed through email attachments (Zaw, 2017). Even though most phishing attacks target the general public, more specific targets such as banks, hospitals, defense organizations, and private companies have been identified as particularly attractive targets. The main motive for these particular attacks is because they have been identified as particularly attractive as they have access to an extensive network of data they collect and hold, including personally identifiable information, confidential information, and financial data (Bose, Leung 2008). In a survey conducted by Wombat Security Technologies, 76% of the 500 information security professionals claimed that their organizations had been victims of a phishing attack in 2016. The consequences of these attacks included compromised accounts and disruption of employees' activities to loss of data (Zaw, 2017). The continuing vulnerability of many organizations to phishing attacks has led the U.K. National Cyber Security Centre recently released specific guidance for organizations regarding how they can defend themselves from phishing threats (NCSC, 2018).

The weakest link in the most secure systems has been suggested to be human error. Existing literature repeatedly identifies peoples' inadequate capacities for detecting online phishing, with more than 90% of individuals falling victim to some form of phishing (Hong, Kelley, Tembe, Murphy-Hill, & Mayhorn, 2013). Participants failed to detect 47% of phishing stimuli and spent little time focusing on security indicators even after initially been primed to do so (Alsharnouby, Alaca, & Chiasson, 2015). Unfortunately, despite the prominence of human risk, research has shown that phishing detection software such as security indicators and toolbars is mostly limited. Studies have shown that people increasingly disregard warnings from security software (Purkait, 2012). Thus, external security systems' limitation highlights the need to investigate human factors concerning cybersecurity and phishing emails. Susceptibility to phishing attacks is referred to as the inability to detect phishing emails and individual differences, which lead to some people being more susceptible than others. This study will be focusing on applying usability methods in order to categorize phishing emails effectively.

Grimes, Hough, and Signorella indicate how little action the user is willing or able to take against the problem (Grimes, Hough, & Signorella, 2007). Some users quickly move illegitimate emails to their junk folder; many do not even take this small step (Williams et al., 2019). This problem arises because of a mental conflict that occurs in the user's mind due to the uncertainty that arises from a phishing email's characteristics.

To combat this arising fraud of phishing emails, we comprised a study consisting of phishing emails with a few prominent features and legitimate emails, which was then presented to participants. Some of these features include awkward prose, wrong spelling, unsolicited attachments, inaccurate imagery, or sending sensitive information via email. We wanted to investigate what kind of correlations a user makes before categorizing an email as phishing or legitimate by conducting this study. The data collected should help shed light on finding a solution to combat this emerging problem.

Characteristic Dimensions of a Phishing Email

<u>Irregular Spelling</u>

Irregular spelling is defined as when a word is spelled differently from what is typically accepted. This a hallmark characteristic that many scammers use in their phishing emails to target potential victims. There are many reasons why scammers choose to add an irregularity in spelling, as many believe the email seems more authentic and believable because, as humans, we are prone to make mistakes. Also, scammers who have minor spelling errors in their emails have a greater chance of their email going through the spam filter than those with proper spelling. Ultimately, scammers only want to target people who will fall prey to these kinds of emails and respond, not have many people respond.

Awkward Prose

Awkward prose is defined as any text that does not fit standard grammatical and lexical forms (Williams et al., 2019). This is another prominent feature that is often seen in phishing emails and is an essential characteristic many users utilize to distinguish phishing emails from legitimate ones. Awkward prose, like irregular spelling, targets users gullible enough to fall for the scam and not an exceptionally broad audience. Spam filters often identify specific grammatical or lexical errors common among phishing emails or search for particular words (Bergholz et al., 2010). Studies conducted also show that these filters are only effective against spam and not phishing emails in particular, even though they share some standard features (Fette, Sadeh, & Tomasic, 2007).

Hyperlinks

A hyperlink is defined as a link from a hypertext file or document to another location or file; these are typically activated by clicking on a highlighted word or image on the screen. Hyperlinks are another characteristic feature that is commonly seen in a phishing email. Users who unknowingly click on such malicious hyperlinks usually lead to various phishing tactics that exploit the user, for example, collecting personal information or downloading malware. Phishing hyperlinks usually contain unsolicited messages or have an embedded URL that has a link different from the link shown.

Use of Imagery

Phishing emails take advantage of brand images, logos, and Q.R. codes to target a potential user. These images are often readily available on the internet to trick users into thinking the email originates from the original company. Phishing emails often contain a signature, including a URL or an underlying Q.R. code. The images usually slightly change the original image with a change in size, color, or tone.

Requesting Sensitive Information

Sensitive information is defined as the data that must be guarded against unauthorized access and unwarranted disclosure to maintain an individual or organization's information security. Requesting sensitive information is the most crucial tactic by which phishing emails scam users. This information can be used to purchase items with the individual's credit card information or even result in identity theft, leading to access and misuse of a user's social networking profile.

Requires a Quick Response - Urgency

Urgency is defined as importance requiring swift action. Various studies show that phishing emails are most effective when they include a component of urgency in the body. Rash decisions are more prone to be made by potential targets or victims when they feel a sense of urgency in complying or doing a specific action as requested in the particular email. The most common words seen in phishing emails that are related to urgency are words like 'notification,' 'immediately,' 'expires,' ‘last final chance or warning’ are all designed in a way to get a quick response.

METHOD

Participants

The participants were students at the University of Central Florida. A total of 20 participants took part in the study and were selected based on their voluntary interests. The demographics collected comprised of age, sex, ethnicity, and education level. The participants' age range was between 18-24, and most of which were students enrolled in the General Psychology course. They received credit for the completion of this study for their course.

Materials

For this study, Qualtrics was used; the materials required were only a working device (phone, laptop, etc.) with an active internet connection. Instructions were shown in a document on the computer monitor before the participants started the study.

Procedure

The procedure was divided into a few parts: (1) informed consent, (2) the experiment, and (3) analysis. The informed consent was given before the participants started the study during their assigned time. Since this study is online, they were presented with a document that contained all the necessary information. After agreeing to the study, the experiment automatically began. The participants were requested to turn off all other devices and not listen to music while partaking in the study to avoid distractions.

The experiment was conducted via Qualtrics, an instruction page was presented on the screen, and the participants were asked to read the directions carefully. Then the participant was presented with 20 phishing emails and 20 legitimate emails in a randomized order. The participant was then asked to categorize whether the email presented him/her was either phishing or legitimate. After choosing what they thought was correct, a text box was presented to write their views on why they categorized it as the option they chose, and their answer is scored on the basis of six dimensions which are listed in the analysis section. The participants were expected to convey their thoughts in no less that than three sentences. The participant was presented with forty questions that contain both legitimate and phishing emails in a randomized order. No prior information will be given to the participant, and they will have to effectively classify them into the best category according to their prior knowledge. The study is approximated to take around 40 minutes.

Analysis

The participants were asked to check for the six dimensions mentioned as follows: irregular spelling, awkward prose, hyperlinks, use of imagery, requesting sensitive information, and requiring a quick response (urgency). The data was scored by counting the number of times each participant mentioned each of the above features in their explanation. The words do not have to be the same but are required to convey a similar idea. We expected that the total number of features mentioned might be related to the subject's phishing email classification accuracy. Prior experience with such exposure may also play a role in accurately classifying phishing emails. Finally, the more experienced a participant is with phishing emails, the more features the participant will mention.

RESULTS

Email Classification Accuracy

Accuracy was examined by email type in order to analyze whether participants could tell the difference between phishing and legitimate emails. Accuracy was measured in terms of the number of correct classifications out of 20 possible ones for both phishing and legitimate emails. As shown in Figure 1, average accuracy was higher for phishing emails (M = 16.93, SD = 2.22) than for the legitimate emails (M = 16.24, SD = 2.89). A paired samples t-test indicated that this was not significant , t(28) = 1.01, p = 0.32, d = 0.19. It appears participants were able to distinguish phishing from legitimate emails well in this study and performed equally for both email types.

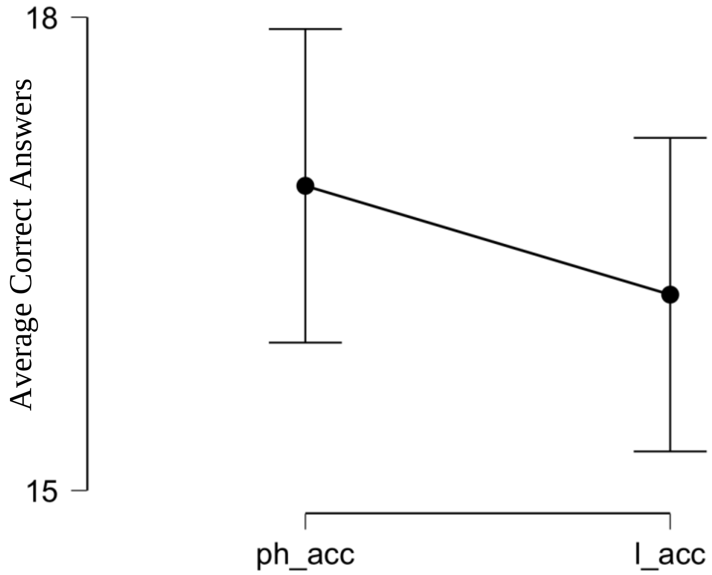

Figure 1: Average Number of Correct Classifications for Phishing and Legitimate Emails

Ph_acc = phishing email accuracy l_acc = legitimate email accuracy

Vertical Axes Represents Average Number Correct out of 20

Email Features Reported

For each email, participants were asked to both classify the email as phishing or legitimate. They were then asked to explain in writing what it was about the email that led them to believe what they did. The answers given by the participants were then analyzed for the presence of six features, namely irregular spelling, awkward prose, hyperlinks, use of imagery, requesting sensitive information, and requires urgency. As shown in Figure 2, the average number of features reported for phishing emails was higher than the total features reported in legitimate emails. A paired samples t-test indicated this was significant, t(28) = 9.454, p < 0.001, d = 1.756.

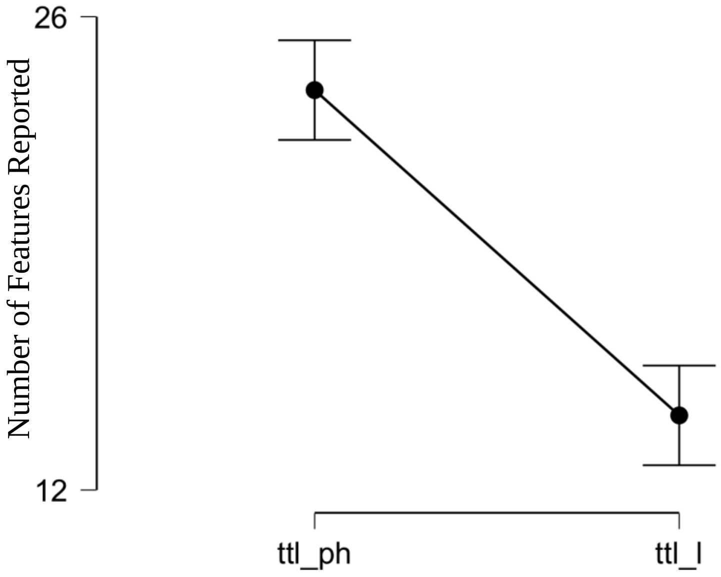

Figure 2: Total Number of Phishing and Legitimate Features Reported

ttl_ph = total phishing ttl_l = total legitimate

Vertical Axes Represents the Number of Features Reported

Features and the Number of Times Mentioned

Figure 3 shows the average number of features reported for phishing emails for each feature type examined. There was a significant main effect of type, F(1,28) = 89.38, p < 0.001, $\eta^2 = 0.052$

There was a significant effect of email feature, F(5, 140) = 17.009, p < 0.001, $\eta^2 = 0.275$

There was a significant interaction between type and features, F(5, 140) = 11.872, p < 0.001, $\eta^2 = 0.061$.

As we can see from Figure 3, there is significant interaction between features and email type. More features were mentioned for phishing emails but are only seen for three dimensions which are irregular spelling, awkward prose, and hyperlinks.

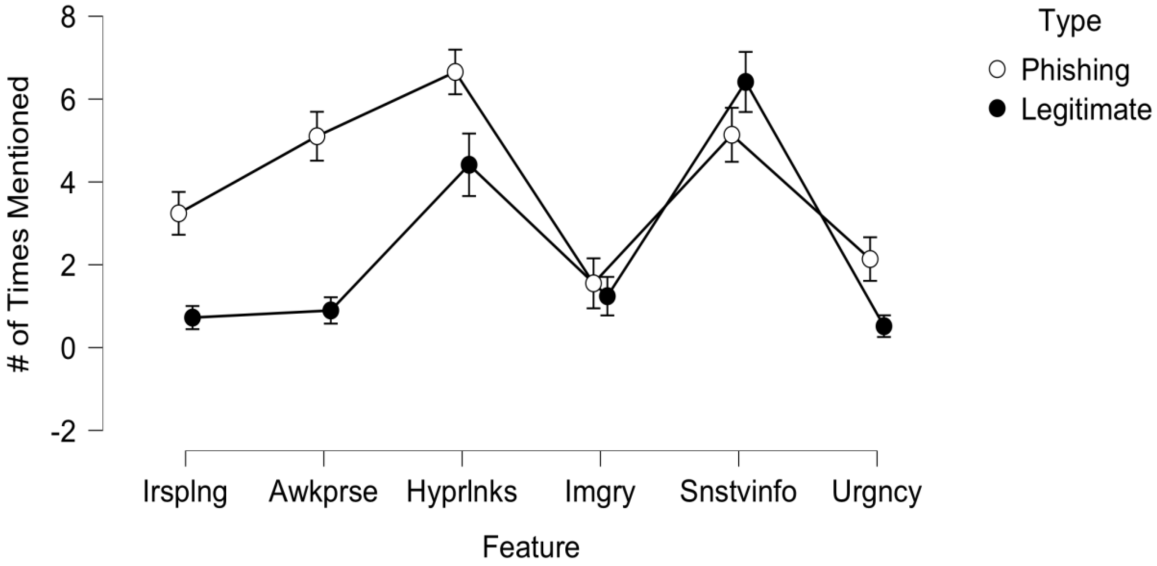

Figure 3: Repeated Measures ANOVA between Features and Number of Times Mentioned

Key
Irsplng - Irregular Spelling
Awkprse - Awkward Prose
Hyprlnnks - Hyperlinks
Imgry - Use of Imagery
Snstvinfo - Requesting Sensitive Information
Urgency - Requiring Urgency

Total Accuracy and Total Features

In Figure 4, the total accuracy was plotted on the x-axis against the total features on the y-axis; a significant positive correlation was predicted to be seen. As it can be seen in Figure 4, there is a significant correlation between total accuracy and total features reported (r = 0.413, p = 0.026). The more features a participant can identify, the higher their accuracy is in distinguishing between phishing and legitimate emails.

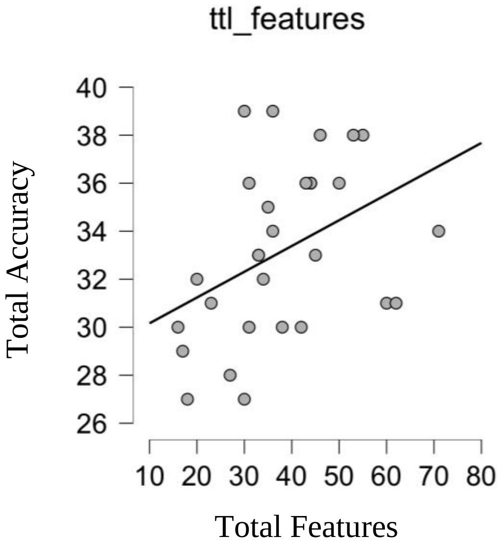

Figure 4: Relationship Between Total Accuracy and Total Features

Vertical Axes Represents the Total Accuracy (x-axis) and The Features (y-axis)

Total Phishing Accuracy and Total Phishing Features

In Figure 5, we predicted that the total number of features mentioned for the phishing emails would predict the phishing emails' accuracy. The x-axis was represented by phishing accuracy, and the y-axis with the total number of phishing features reported. There appears to be little to no correlation between the two variables, Pearson's correlation showed a slight correlation between the number of features listed (r = 0.186, p = 0.333). It does not appear that phishing emails' accuracy was related to the number of features mentioned in this study.

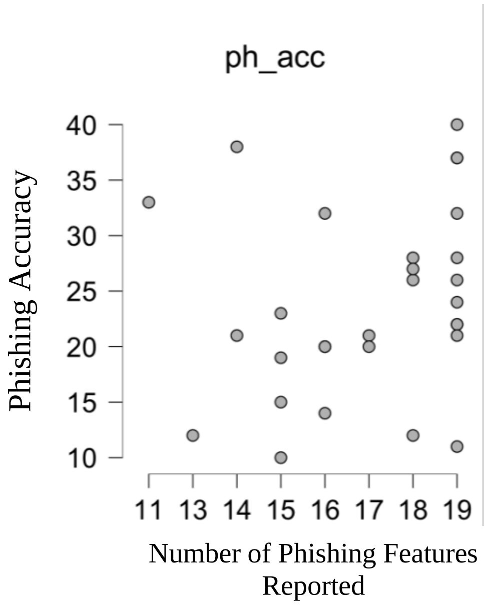

Figure 5: Relationship between Total Phishing Features and Phishing Accuracy

Vertical Axes Represents the Phishing Accuracy (x-axis) and The Number of Phishing Features Reported (y-axis)

DISCUSSION

This study examined how accurately a participant could categorize between phishing and legitimate emails using prior experience. Participants were asked to categorize twenty phishing and twenty legitimate emails, which were presented to them in a random order, into the mentioned categories; an explanation on why they chose the category they did was also required. We formulated the hypothesis that the more features they said while classifying an email as phishing or legitimate, the better they would be at classification. Hence, a strong positive correlation would be noticed. Correlation analysis, repeated measures ANOVA and t-tests were conducted on the data; the results obtained were not on par with our hypothesis.

When analyzing the results of the t-tests, we could see that participants could accurately differentiate between phishing and legitimate emails. Also, participants were able to point out more features in phishing emails than legitimate emails. The repeated measures ANOVA showed us there was a significant main effect on the type and a substantial effect of the email features. It was also statistically proved that there is a significant interaction between type and features in this study.

However, the correlation analysis showed little to no correlation between the number of features mentioned by the participant and accurately classify an email as phishing, which disproves our hypothesis. However, there was seen to be a strong correlation between the total features mentioned and the total accuracy, and if a participant mentions more features, there is a bigger chance of making a correct guess about whether the email was phishing or legitimate. So, despite having vast exposure to phishing emails, it doesn't necessarily correlate with the fact that the person can pinpoint the features effectively and not fall into the traps such emails set for innocent and unsuspecting victims.

Limitations of the Study

The present study didn't ask for the participant for their level of exposure to such phishing emails, so if subsequent studies can quantify, it would be beneficial to take studies related to this further. Also, only one researcher analyzed the data collected from the administered survey so that researcher bias may be present; to avoid this, if an additional one or two research could also pitch in, it may lead to much more reliable results.

Additional Studies that can be Conducted

Additional studies that can be conducted on the same topic may include more cognitive tasks that help us understand what factors are the ones that cause victims to fall into the trap of such phishing emails and how to prevent them. Also, for a subsequent study that can be conducted, we could utilize the limitation discussed previously, the quantification of exposure level, to analyze the correlation between that and accurate categorization of phishing emails to get more reliable data. Results from these studies can help us combat the growing issue of cybercrime and cyber phishing in today's world.

How to Combat Phishing in the Healthcare Sector

Another growing concern about phishing crime is that of the one in healthcare. The crimes increase in number and sophistication each passing day. This mainly arises due to improper large-scale protection offered in the health sector. Cyber attackers can do various fraudulent things if unprotected health information gets into their hands, making a cash in on the providers' money and data. This is another grave aspect of phishing crime that tends to go overlooked and needs to be tackled.

Various methods can be undertaken to tackle phishing crimes, specifically in the healthcare sector, starting with training the health sector workforce on how to tackle incidents relating to cybercrime. For this, various training and educational sessions have to be conducted for the employees to learn and acquire skills on how to deal with such situations. Secondly, the organizations that are prominent in the healthcare field should find a way to minimize the amount of information available to the public. Thirdly, filtering out questionable content is a given in this case as well having a multifactor authentication works very well in tackling this problem.

How to Combat Phishing in the Tax Sector

Another cyber phishing sector I want to shed light on is the rising ones dealing with tax identity theft. These phishing scams aim to gather sensitive data that includes but is not limited to passwords or account information through emails, texts, or weird websites. An important point to note is that IRS will never ask or contact a taxpayer through any cyber platform like social media, text messages, or emails for any financial or personal information.

Various strategies could be undertaken to prevent fraudulent activity from IRS imposters and identity thieves trying to steal your personal information. The most important step is by protecting one's social security number (SSN); this is something that should not be given to anybody unless there is a good reason to do so. Filing your tax return as soon as the tax season starts and using a strong internet connection to do so are good ways to prevent fraud. A thorough check should be done on a tax preparer if you are choosing to hire one, and lastly, checking your credit card report at least once a year to make sure no new accounts were opened up in your name.

Educating the General Population about Phishing Scams

Another aspect I would like to stress on is educating our population about such scamming and phishing activities. It is no surprise that the most scammed using such phishing crimes are the elderly. This may be due to them being more frequent users of emails and are more likely to fall into the trap of a phishing email than their younger counterparts. Ultimately, I want to stress that educating the young and old alike on the importance of not falling victims to such phishing attacks is of grave importance. Educating the young at schools would be a great place to start, and holding phishing awareness camps for adults would be another great idea. By implementing such small strategies, I feel like win over phishing attackers and end phishing one day for good.

REFERENCES

Alsharnouby, M., Alaca, F., & Chiasson, S. (2015). Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human-Computer Studies, 82*, 69-82. doi:10.1016/j.ijhcs.2015.05.005

Bergholz, A., De Beer, J., Glahn, S., Moens, M., Paaß, G., & Strobel, S. (2010). New Filtering Approaches for Phishing Email. *Journal of Computer Security, 18*(1), 7-35. doi:10.3233/JCS-2010-0371

Bose, I., & Leung, A. C. (2008). Assessment of Phishing Announcements on Market Value of Firms. *2008 International Conference on Information Technology*. doi:10.1109/icit.2008.37

Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to detect phishing emails. *Proceedings of the 16th International Conference on World Wide Web - WWW '07.* doi:10.1145/1242572.1242660

Gordon, W. (2017). Threats to Information Security - Public Health Implications. *The New England Journal of Medicine, 377*(8), 707-709. doi:10.1056/NEJMp1707212

Grimes, G. A., Hough, M. G., & Signorella, M. L. (2007). Email end users and spam: Relations of gender and age group to attitudes and actions. *Computers in Human Behavior, 23*(1), 318-332. doi:10.1016/j.chb.2004.10.015

Hong, K. W., Kelley, C. M., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2013). Keeping Up With The Joneses. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 1012-1016. doi:10.1177/1541931213571226

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, *50*(10), 94-100. doi:10.1145/1290958.1290968

Leyden, J. V. (2004). You Just Don't get It! *Index on Censorship, 33*(2), 101-105. doi:10.1080/03064220408537336

Phishing attacks: Defending your organisation. (n.d.). Retrieved November 19, 2020, from https://www.ncsc.gov.uk/guidance/phishing

Purkait, S. (2012). Phishing Counter Measures and their Effectiveness- Literature Review. *Information Management & Computer Security, 20*(5).

Van Dyke, J. (2004). Online Account Management as the Antidote to Fraud: Financial Institutions and Billers Must Revamp Their Web Features and Messages. *Javelin Strategy & Research*.

Williams, S. E., Sarno, D. M., Lewis, J. E., Shoss, M. K., Neider, M. B., & Bohil, C. J. (2019). The psychological interaction of spam email features. *Ergonomics*, *62*(8), 983-994. doi:10.1080/00140139.2019.1614681

Zaw, T. (n.d.). Blog: 2017 Verizon Data Breach Investigations Report (DBIR) from the Perspective of Exterior Security Perimeter: Verizon Media Platform. Retrieved November 19, 2020, from https://www.verizondigitalmedia.com/blog/2017-verizon-data-breach-investigations-report/