Comparing Scam Emails and Email User Education at Universities

Duo Pan Ellen Poplavska Nora O'Toole Shomir Wilson
Pennsylvania State University

Abstract

Scam emails pose a threat to the personal information and financial safety of all email users, and user education is a common response to the proliferation of email scams. However, existing research exploring how well scam education represents actual observed scam emails is limited. In particular, existing research has not compared the current landscape of user education regarding scam emails to a dataset of genuine scam emails, to determine similarities and differences between educational materials and actual scams received by email users at organizations that have the resources and incentives to create these materials. We examine this gap using metadata of emails sent to a scam email reporting address at the Pennsylvania State University (*Penn State*), a large research institution in the United States. We compare this dataset with the scam email user education materials at 16 peer universities, observing differences between scams that users receive and security training materials their institutions provide.

1 Introduction

Scam emails, which can lead to identity theft and financial loss, are increasing in number. According to the annual State of the Phish Report, users reported nearly 9.2 million suspicious emails in 2019, which is 67% higher than in 2018 [8].

According to the same report, 51% of US workers cannot recognize the concept of "phishing" correctly [8]. Without recognizing the concept of phishing, users may not know how to recognize these malicious emails, and what actions, such as reporting a scam, they may take when they receive them. According to Purdue University's report in 2018, 61% of emails sent to purdue.edu email addresses were malicious. However, only 1,711 people reported potential scam emails to the university [10]. It is likely, therefore, that many Purdue email users received malicious emails and did not report such emails.

User education is one method that large organizations use to encourage users to be mindful of security. To date, however, there has been no work comparing this user education to the landscape of scam emails that email users within such an organization face. In order to bridge this gap, we examine a dataset of the subject lines of scam emails that had been sent to the Office of Information Security (*OIS*) at Penn State University between 2020/08/12 and 2020/11/11. OIS collects malicious emails forwarded from users of Penn State's email system, including students, faculty, and staff at the university. We compare this dataset to the scam email user education resources provided by the 14 universities in the Big Ten Conference and its two affiliate member institutions, and observe mismatches between the types of emails prevalent in the dataset and those prevalent in the resources [11]. However, one limitation to this is that the type of scam emails Penn State's OIS receives may not be representative of the emails other Big 10 universities receive.

2 Related Work

Prior work has examined the role of user education in cybersecurity. Tokata and Ogura proposed that user education could be useful in combating targeted attacks [9]. They first explored the relationship between human psychological characteristics and vulnerability against social engineering. Testing these characteristics can predict whether the user is vulnerable to a specific social engineering technique. They then developed web-based learning materials to counter these techniques.

Harley and Lee [3] presented a study of the landscape of current scam email education. They reviewed a range of web-based educational and informational resources and analyzed

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021.
August 8–10, 2021, Vancouver, B.C., Canada.

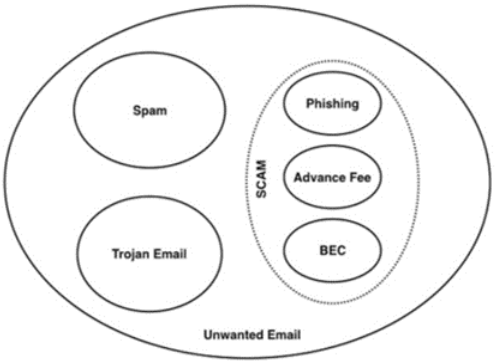

Figure 1: Jakobsson's classification of unwanted email [5]

the benefits and drawbacks of quiz-based learning modules. These quizzes were found to be engaging and interactive, but presented a simple view of the scam emails that only covered one or two of their characteristics.

Pawar and Tijara [7] introduced various methods of user scam education. For instance, the PhishGuru coaching system helps users identify scam attacks. They observed that current scam email user education is not effective because users do not understand how scammers operate, and instead make assumptions based on their own real-world experiences. Therefore, effective user education should allow users to correctly assess potential risks and benefits, rather than only warning them to beware of danger.

In Figure 1, Markus Jakobsson [5] indicated the difference between scam, spam, and phishing emails. Within all unwanted emails, spam emails are legitimate emails designed to sell products and services. In contrast, scam emails deceive users for financial gain, and phishing emails are a type of scam used to steal users' credentials.

Finally, to our knowledge, little prior work has examined gaps between common assumptions about email scams and actual specimens in the wild. Pan et al. [6] showed thematic differences between collections of scam emails in multiple languages, raising concerns that English-centric training may not represent all email users' experiences. In contrast, we focus on a gap between scam emails observed in an English-speaking setting and the examples shown to users in educational materials.

3 Data Sources

3.1 Email Metadata

OIS shared with the researchers the metadata of 2,092 emails. These emails were forwarded to an email address designated to collect and analyze malicious emails within the Penn State community.

The metadata contains three attributes for each email:

- Time: The time that OIS received the email, ranging from 2020/08/12 to 2020/11/11
- Subject: The subject line of the reported emails
- Tag: The annotated tag to indicate the scam email category

OIS then annotated each reported email with one of 23 tags to describe the type of email. 2084 emails receive only one tag, and eight of them received two tags. These tags sort the emails into categories of similar specimens. OIS annotated the emails by hand, and added new tags in cases where existing tags did not adequately describe an email.

We list the 10 most frequent tags and brief descriptions of each in Table 1. Table 2 shows each tag's frequency, the percentage of the dataset annotated with this tag, and an example subject line.

3.2 Big Ten Universities' User Education

Based on these findings, we analyze how universities' email user education compares to the particular problems that scam email recipients at educational institutions face. We examine resources from the 16 universities (14 members and two associate members) that belong to the Big Ten Conference [11] as an example. We chose the Big Ten Conference due to the similarities between its member institutions, as well as the relatively large sizes of these institutions and their resources, which lead us to expect that they are able to provide user education regarding scam emails. Additionally, because our scam email subject line dataset was obtained from Penn State, a Big Ten university, this sample best allows us to generalize observations about our dataset to this entire set of similar institutions.

We pose three questions about each universities' scam email user education:

- Which types of scam emails does this user education address?
- Does the university's user education differentiate scam emails, spam emails, and legitimate emails?
- Does the university's user education specifically address university-targeted scam emails?

Researchers used Google [1], a search engine, to find the user education websites provided by the 16 universities as a source of gathering the data and coded, by hand, the schemes presented below. The term "phishing education" with the university's name was typed into the search engine and the first link found was used for collecting the data.

Tag	Description
Phishing-Gift-card	An email that asks the recipient to purchase a gift card
Phishing-General	A phishing email that does not fit another category
Phishing-No-Response	An email that OIS will not respond to
Phishing-Ube	An unsolicited bulk email (*UBE*)
Phishing-Malware	An email containing malware or a link to a site hosting malware
Phishing-General-Scam-Extortion	An email using threats to obtain information
Phishing-Legit-Email	A legitimate email mistakenly identified as a scam by the person forwarding the email
Phishing-Job-Offer	A scam email containing a false job offer
Phishing-Request-Resend	An email that OIS requests the receiver to resubmit as an attachment
Phishing-Ms-Quarantine	A legitimate email from Office 365 notifying the user about a quarantined malware email

Table 1: The 10 most frequent tags in the OIS dataset accompanied by a brief description of each

Tag	Frequency	Percent	Example Subject Line
Phishing-Gift-Card	486	23.2%	Got a moment
Phishing-General	483	23.1%	Private Message
Phishing-No-Response	223	10.7%	Your daily briefing
Phishing-Ube	220	10.5%	MagScoop is the BEST Scooper for your Kitchen!
Phishing-Malware	217	10.4%	eMail Security Check!
Phishing-General-Scam-Extortion	148	8.1%	Diplomat Arrival
Phishing-Legit-Email	112	5.4%	Penn State University - Clearance Information Needed
Phishing-Job-Offer	96	4.6%	BABY SITTING
Phishing-Request-Resend	34	1.6%	Phishing Attempt - Teams Email
Phishing-Ms-Quarantine	23	1.1%	Spam Notification: 1 New Messages

Table 2: The 10 most frequent tags in the OIS dataset and their frequency, percentage of the dataset, and example subject line

University Name	Legitimate vs. Scam	Spam vs. Scam	University-Targeted Email
Indiana University	•		•
University of Maryland	•		•
University of Michigan	•		•
Michigan State University	•		
Ohio State University	•		•
Pennsylvania State University	•	•	•
Rutgers University	•		
University of Illinois	•		•
University of Iowa	•	•	•
University of Minnesota	•	•	•
University of Nebraska	•	•	•
Northwestern University	•	•	•
Purdue University	•	•	•
University of Wisconsin	•	•	•
Johns Hopkins University	•	•	•
University of Notre Dame	•		

Table 3: Topics represented by descriptions or examples in Big Ten Universities' scam email user education

4 Analysis

4.1 Observations on the User Resources

To answer these three questions, we first analyze the user education resources provided by Big Ten universities.

Scam email examples provided by universities represented a number of different types of scams; however, the distribution of these different scam types differed substantially from the distribution of scam types observed in the OIS dataset. The OIS dataset contained more gift card scams than the examples observed in the educational resources. We find that all analyzed universities define a scam email and explain how it is different from a legitimate email, but eight of 16 universities fail to explain the differences between spam and scam emails. 13 of 16 universities also noted that scam emails can target university email users. However, three universities did not describe scams that target university email users specifically (such as pension scams targeting faculty).

Frequent updates to user resources provide email users with pertinent information about scam trends; therefore, we observe the frequency with which these universities update their scam information resources. The University of Michigan updates its scam email examples every month [4], but most Big Ten universities update their examples less frequently. Although some universities did not update their scam email examples frequently, all surveyed universities note the presence of scams related to the COVID-19 pandemic, indicating that these universities have revised their educational materials within the last few years to account for this event and the new scam emails relating to it.

4.2 Observations on the OIS Dataset

We observe some pertinent features of the OIS dataset. The most common scam emails that Penn State email users receive are Phishing-Gift-Card, Phishing-General, and Phishing-Unsolicited Bulk Email (*Ube*). 10.5% of reported emails are tagged as spam emails. This suggests that email users had trouble differentiating between deceptive scams and truthful, albeit unwanted, spam emails.

Some users received email subject lines targeting their institution, such as those claiming to provide information about pension plans at Penn State University. Because university-targeted scams are present, educational resources provided by universities ought to address or represent them.

4.3 Comparisons Between the User Resources and the OIS Dataset

Table 3 shows the results of applying each of these three questions to each university's publicly accessible online resources. We do not include any resources that require a login to access.

First, we compare the categories of scam emails in the OIS dataset and universities' scam email education. For instance, within the OIS dataset, email users are most likely to receive gift card scam emails, but 11 of the 16 universities do not address this scam type in their user education resources. 10 of the 16 universities provide an example of a job offer scam, but only 4.6% of the OIS dataset consists of job offer scam emails. University educational materials therefore place greater emphasis on job offer scams than the apparent frequency of these scams among scam emails. In contrast with job offer scams, other types of scams are drastically underrepresented.

Next, we investigate whether universities instruct their email users to differentiate scam emails from legitimate emails and spam emails. In Table 3, we find that all 16 universities provide some strategies for differentiating between legitimate and scam emails. However, only eight universities explain differences between spam and scam emails.

Third, we explore whether these 16 universities explain that some scam emails target specific universities. In Table 3, we notice that 13 universities mention university-targeting scam emails in their user education materials. Three of them indicate that some scam emails target international students. For instance, The Ohio State University listed several types of scam emails that affect international students, including immigration scams, tuition scams, travel scams, and package and mail scams [2].

5 Discussion and Future Work

After comparing the OIS dataset with 16 universities' user education, we find that most of these universities accurately define a scam email, providing a basic level of guidance to email users regarding email security. However, there is a divide between the types of scam emails in the OIS dataset and universities' user education. Gift card scams are underrepresented in the educational resources.

Almost half of these universities do not explain the differences between scam and spam emails. This is an oversight that may leave users without guidance in differentiating malicious emails and mass mailings.

Most universities mentioned that some scam emails specifically target university users, such as faculty or international students, indicating that this type of scam is well-represented within university resources.

This work observes a gap between the types of scam emails represented in scam education resources and those in user's inboxes. This gap suggests that changes to these resources may provide a more accurate representation of email users' experiences with email scams. Further research might investigate potential similar gaps between other groups of email users outside of higher education.

Acknowledgments

This work was supported in part by a grant from the Center for Security Research and Education at Penn State University. The email metadata described in this paper is provided by Office of Information Security at Penn State University, with thanks to Chris Decker, Richard Sparrow, and Holly Swires.

References

[1] Google. https://www.google.com, 2021.

[2] Student Legal Service at the Ohio State University. International Students Scam. https://studentlegal.osu.edu/international-students/scams/, 2021.

[3] David Harley and Andrew Lee. Phish phodder: Is user education helping or hindering? In *Virus Bulletin Conference Proceedings*, pages 1–7, 2007.

[4] Information and Technology Service at University of Michigan. Phishes Scams. https://safecomputing.umich.edu/phishing-alerts/, 2021.

[5] Markus Jakobsson. *Understanding social engineering based scams*. Springer, 2016.

[6] Duo Pan, Ellen Poplavska, Yichen Yu, Susan Strauss, and Shomir Wilson. A multilingual comparison of email scams. 2020.

[7] Vishakha B Pawar and Pritish A Tijare. User security awareness against phishing. *International Journal*, 2(3), 2014.

[8] Proofpoint. State of the phish. pages 1–48, 2020.

[9] Toyoo Takata and Kanayo Ogura. Confront phishing attacks—from a perspective of security education. In *2019 IEEE 10th international conference on awareness science and technology (iCAST)*, pages 1–4. IEEE, 2019.

[10] Purdue University Information Technology. Email scams and phishing – how to spot them in your Purdue email and what to do. https://www.itap.purdue.edu/newsroom/2019/190405-How-To-Spot-Phishing.html, 2021.

[11] Big Ten Conference Official Website. Big Ten Conference. https://bigten.org/, 2014.