An Analysis of Phishing Reporting Activity in a Bank

Doing, Anne Kee; Barbaro, Eduardo; van der Roest, Frank; van Gelder, Pieter; Zhauniarovich, Yury; Parkin, Simon

DOI
10.1145/3688459.3688481

Publication date
2024

Document Version
Final published version

Published in
Proceedings of the 2024 European Symposium on Usable Security, EuroUSEC 2024

Citation (APA)
Doing, A. K., Barbaro, E., van der Roest, F., van Gelder, P., Zhauniarovich, Y., & Parkin, S. (2024). An Analysis of Phishing Reporting Activity in a Bank. In *Proceedings of the 2024 European Symposium on Usable Security, EuroUSEC 2024* (pp. 44-57). (ACM International Conference Proceeding Series). Association for Computing Machinery (ACM). https://doi.org/10.1145/3688459.3688481

An Analysis of Phishing Reporting Activity in a Bank

Anne-Kee Doing
Delft University of Technology
Netherlands

Eduardo Barbaro
ING Bank and Delft University of Technology
Netherlands

Frank van der Roest
ING Bank
Netherlands

Pieter van Gelder
Delft University of Technology
Netherlands

Yury Zhauniarovich
Delft University of Technology
Netherlands

Simon Parkin
Delft University of Technology
Netherlands

Abstract

A reduction in phishing threats is of increasing importance to organizations. One part of this effort is to provide training to employees, so that they are able to identify and avoid phishing emails. Yet further, simulated phishing emails are used to test whether employees will both identify and report a suspicious email. We worked with a partner bank to examine a repository of many thousands of reported emails from a behavioural perspective. We divide reported emails into categories and examine reporting trends over time relative to training and phishing simulation campaigns. Among our findings, the level of reporting of benign emails is comparable to the number of malicious emails reported, and we see indications that training and simulations amplify the reporting of benign emails. Our analysis uncovers reporting patterns for unique reporters per email campaign as a promising indicator for the security-related culture around phishing prevention. Evidence from our analysis informs recommendations, such as providing reporting infrastructure for reporting not only malicious emails, but also benign but suspicious work-related emails, in a manner that minimises the disruption for users erring on the side of caution when assessing emails.

CCS Concepts

• **Security and privacy** → **Usability in security and privacy**; **Phishing**.

Keywords

Phishing reporting, user email reporting, phishing simulations

ACM Reference Format:
Anne-Kee Doing, Eduardo Barbaro, Frank van der Roest, Pieter van Gelder, Yury Zhauniarovich, and Simon Parkin. 2024. An Analysis of Phishing Reporting Activity in a Bank. In *The 2024 European Symposium on Usable Security (EuroUSEC 2024), September 30–October 01, 2024, Karlstad, Sweden.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3688459.3688481

EuroUSEC 2024, September 30–October 01, 2024, Karlstad, Sweden
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1796-3/24/09
https://doi.org/10.1145/3688459.3688481

1 Introduction

As a countermeasure to phishing attacks on organizations, email spam filters and phishing detection methods are improving [27], but some malicious emails will still arrive in an employee's inbox. This has translated to challenges for employees to understand cybersecurity and manage phishing emails, toward being able to work securely; outcomes of staff behaviour can increase or decrease companies' cybersecurity drastically [28]. One of the approaches to increase security awareness among staff is to provide security training for employees [2]. This leads to potential for phishing emails not detected by technological solutions to be identified and reported by employees [12, 35].

It is rare that an aspect of security-related training is tested in practice [23]. This research investigates which lessons can be learned from the reported emails, by conducting exploratory research on a case study in a bank. This leads to our main research question: *How can email reporting patterns in a large organisation measure the relationship between phishing training, reported emails and employee behaviour?*

This research studies the reporting behaviour of employees at a bank with approximately 60,000 employees, at a large scale through 50,000 reported emails; this involves thousands of reported emails per week, over the period from 1st January 2022 to April 1st 2023, covering a period of 16 months. Within this time period, employees underwent E-learning, but also wide-scale phishing simulation waves that were deployed by the bank.

With the reported emails in a bank, a characterisation of the behaviour of employees is constructed. We identify trends by visualising reported email counts and several attributes over time. Furthermore, with Natural Language Processing (NLP) methods, the emails can be compared based on the content of the reported emails. The comparison between the content of the simulation emails and the reported emails can provide valuable insight, into the effect of the simulation on the employees and their reporting behaviour. Additionally, extracting email topics affords some comparison of attackers' tactics and suspicious-looking but benign emails. With the NLP technique TF-IDF, the content of the emails is compared, and the dominant topics are found.

As a contribution, this work documents actual reported emails by active users in a real-world organization. The work also makes a novel contribution in acknowledging that reporting is not always precisely about malicious emails, and that 'benign', non-phishing emails may be reported too – just as phishing emails aim to look like real emails, real emails may inadvertently appear to be like phishing emails. We also use the reporting data to explore potential metrics at scale; this points to the contribution of the method as a means to analyze reported emails at scale and compare their characteristics. Further to this, we compare the presence of particular 'components' across benign and malicious emails, and in comparison to simulated phishing emails delivered by the partner organization to employees. This contrasts with complementary work such as the US NIST Phish Scale [40], which works to identify email components to compare malicious and benign emails, but 'in the small', email by email.

We find that over the analyzed time period, the level of reporting of benign emails is comparable to the number of malicious emails reported. Also, where several hundred benign and known-malicious emails may be reported on a week-by-week basis, widespread reporting behaviour also captures several emails per week which require further investigation. This represents where user reporting does capture threats that technology alone may not necessarily find, relative to known threats and false positives. We also uncover metrics for unique reporters per email campaign, as a promising metric for the culture around reporting and security. The analysis informs recommendations, such as acknowledging that reporting impacts *suspicious* emails (which includes both the true positives of malicious emails, and benign false positives), due to the very same overlap that social engineering attacks aim to exploit. Given the comparable rate of benign reports to malicious reports, we also focus recommendations on ensuring that the reporting process is rapid, including confirmation of the outcome of reporting an email, so that employees can return to productive work if a reported email turns out to be genuine.

Related Work follows in the next section, accompanied by details of the context of the partner organization (section 3). In section 4 we detail our Methodology, and in section 5 the Results of our analysis. We bring the paper to a close with Discussion of the implications of our analysis in section 6 and Conclusions in section 7.

2 Related Work

There are multiple ways to provide phishing training, such as sending fake phishing emails, or a lecture setting where an instructor informs staff about how to recognise phishing emails. Several studies suggest such training can increase security [19, 30], but they are still recent and limited. Generally, training effectiveness is measured with click-rates (the percentage of people who click on a link in the simulation email), where low rates are seen as more secure.

Considering this and the lack of consensus in the literature for the best anti-phishing training [26, 35], the correct way to deal with phishing vulnerability is difficult to determine. However, as stated by [22], a clicking rate of zero is impossible. Some cues can be taught, but no amount of training will bring the clicks to zero [6], and the cost-benefit balance must also be considered here. Anti-phishing measures can potentially disrupt normal working practices [10].

For several banks, phishing is found to be the most common way for attackers to gain access to their systems or networks [17, 25]. Vital sectors such as the banking sector are highly regulated, as a security breach in these sectors could have disastrous consequences for the economy and safety of citizens [34]. To set a base level of security between banks, governments have introduced regulations all banks must adhere to [18].

A commonly used tool to measure employees' security behaviour regarding phishing is a phishing simulation, where companies send fabricated phishing emails to their employees to measure how they interact with them [24, 35]. Correspondingly, some examples of research on phishing reporting focus on the reporting rates of emails by creating a test environment (e.g., [35, 37]). Other approaches include sending simulation emails whereby the participant is in a real-world environment [42], or, having participants answer a questionnaire about their behaviour in real-world circumstances [3]. Additionally, embedded phishing training is not necessarily without its own drawbacks, as the training may make employees more susceptible to phishing, rather than less (e.g., [12, 35]), requiring further examination of the costs and benefits.

The NIST phish scale aims to articulate the difficulty of identifying a phishing email [40], and shows that premise alignment with an employee's job can make a phishing email more difficult to identify correctly. This can help to explain high click rates, and provide an alternative explanation for behaviour besides the characteristics of employees. For example, a benign email may look suspicious and be reported by an employee; rating emails based on characteristics instead of intent makes it clearer where malicious and benign emails overlap and differ. Here we analyze reported emails in a real setting, including malicious and benign emails, to consider further costs that arise from this overlap, in terms of true (malicious) and false (benign) reports.

3 Organization Context

When researching phishing susceptibility, a case study is a frequently used approach, either by sending counterfeit phishing emails to unknowing employees or letting participants complete a survey [20, 21, 38]. By contrast, real phishing emails are rarely used to study employee behaviour. By researching interaction with actual emails in real-world situations, the organisational aspects of susceptibility can be investigated instead of solely the individual's susceptibility [24]. Compared to an experimental setup, where isolating the factors you want to study is possible, a non-controlled setting is more challenging. Namely, external factors cannot be excluded in a non-controlled setting where participants are observed and studied within their everyday environments.

From the 2022 annual reports of several banks in the Netherlands, phishing is the most common way for attackers to gain access to the bank's system [17, 25]. This research focuses on these last forms of attacks aimed at employees and the threats related to breaches on this side. Previous research found that bank employees had better information security awareness than the general public [38]. This suggests their response to emails is also different from that of the general public, creating a subgroup of potential victims with their own security profile.

3.1 Email Flow

When an email is sent, there are multiple stages it can go through to determine whether the email is safe. Once an email arrives in an employee's inbox, they decide how they interact with the email. Most commonly, the email is part of normal email traffic, and the employee responds to the email. In some cases, however, the email appears suspicious to the employee, and they can report the email.

They can do this from within their email carrier by forwarding the email to the correct email address or clicking a 'report phish' button in their inbox. In the last case, the email is automatically forwarded to the correct email address. The email is also immediately moved to the bin folder of the employee's email.

If an employee reports an email, an external company evaluates the email and determines whether it was a phishing attempt or a legitimate one. The final classification is then sent back to the employee. Additionally, if the email is classified as malicious, this is communicated with the department responsible for the email filters, after which they can adapt their anti-phishing filters.

3.2 Training in the Company

Each employee in the company receives various types of training to achieve different objectives, such as creating a safe workspace and promoting responsible banking. All employees also receive anti-phishing training to make employees more resilient against phishing emails. To help employees identify and report suspicious emails, they receive training organised within the company. This is a mandatory E-learning training every employee has to complete upon joining the bank. This knowledge learned through training does have to be maintained. Within the company, the mandatory E-learning training is repeated for all employees periodically, approximately once every year.

The second type of training adopted in the company is sending simulated phishing emails in a wave to all employees. With this approach, employees get an email with a link, and if they click on the link and submit their password, they are sent to a page with information about the fact that the email was part of a test and how they could have recognised it. The cues for identifying the emails as suspicious are taught in the E-learning. The simulation wave is a tool to both measure the behaviour of the employees, and to provide additional selective training to employees who did not recognise the phishing attempt. Not all employees receive these emails, and during the most recent wave, about 50% of the employees were randomly selected. In addition, some departments within the company have embedded simulation training throughout the year, targeting all their employees.

If a reported email is indeed phishing, they receive the feedback they correctly identified the email. If the email was benign, they also get this feedback, signalling that the email was legitimate. If an email is reported, it is transferred to the deleted mailbox. If an email turns out benign, the employee would have to return to their bin folder to retrieve the email. If the benign email contained an action the employee should perform, there are more related costs because the action is postponed. Additionally, the costs of losing employees' trust can be found in the increased resistance towards online communication and safety measures [42].

4 Methodology

Here we describe our research design. Firstly, we describe the dataset and how it was prepared for analysis. From subsection 4.4 onwards, we describe the specific qualities we measured from the dataset. This includes how content was treated in order to compare text across reported emails (malicious, benign, and simulation emails).

4.1 Data Contents

During 2022, employees received E-learning training, and there was a company-wide simulation wave where counterfeit phishing emails were sent to 50% of the approximately 60,000 employees. The content of the emails and the click-, compromise- and reporting-rates of the simulation wave are known. Additionally, the reported simulation emails were part of the larger dataset with all reported emails. A list of all factors present in the dataset about the reported emails can be found in the Appendix (Data Contents). Privacy regulations prohibit using individual characteristics such as age.

Second, one dataset contained information about the mandatory E-learning employees followed. This data includes the course, the moment an employee finished the training and the employee's job description. Due to privacy regulations, the E-learning cannot be traced back to the original employee who finished the training. However, the vast majority of employees finished the training before data for this research was collected. Therefore, as it is impossible to link the E-learning completion to the individual employee, only the aggregated counts were used.

Third, the information about individual employees is contained in another dataset. For each employee, the email address and job description are known. The lowest collected level of information is to which department an employee belongs. Fourth and last, information about the counterfeit emails from the simulation wave is known. Within the wave, ten simulation emails were used. The content of these emails, compromise rates, click rates and number of correctly reported emails are known. By combining these datasets, it is possible to determine when the training took place and the changes in reports over time. We elaborate further on the content of the datasets in subsection 4.3. Once the data was collected and merged accordingly, it was cleaned.

4.2 Data Cleaning

There are some missing values in the employees' job descriptions. Additional information for incomplete entries was provided based on additional data, as described in [13]. By combining additional sources of information and personal communications with company experts, gaps in employees' business lines were filled whenever possible. Unfortunately, this was not the case for all entries. Some emails were classified incorrectly while they were part of the simulation wave. These emails have been found based on their title and given the correct classification. Additionally, it was checked that the sender was as expected. As the number of emails was very narrow (less than 1% of the simulation emails) and the title of the email was known, using only the title to locate these emails was sufficient in this specific case.

4.3 Data Statistics

The data contains 50,000 reported emails with several attributes. The most important attributes which were used for the analysis are outlined below. The dataset contained all emails employees have reported between January 1st 2022 and April 1st 2023. The emails have been categorised into five distinct classes, which are explained below.

- **No Threat Detected**: The email does not contain a threat, and normal interaction with the email by the employee is safe.

Table 1: All classifications and the percentage of emails classified as such.

Classification	Percentage
No Threat Detected	34.8 %
Malicious	26.1 %
Simulation	23.4 %
Do Not Engage	13.3 %
Further Investigation Required	2.5 %

The employee incorrectly identified the email as potential phishing. For brevity, we also refer to these emails as benign emails.

- **Malicious:** The email contains a threat and is correctly identified as potential phishing by the employee.
- **Simulation:** The email is part of a simulation sent from within the company. The employee correctly identified the email as potential phishing.
- **Do Not Engage:** Although the email does not contain a direct threat, the motivation behind the email does not seem legitimate. Therefore the advice is to not interact with the email further. The employee may then be regarded as exercising appropriate caution, given that the email seems malicious or manipulative.
- **Further Investigation Required:** The report does not include enough information to classify the email. The main reason for this is an incorrect method for reporting the email, for example, without the suspicious email body or title attached, making it impossible to classify the email and determine whether it is a threat.

The percentages of the classifications in the dataset can be found in Table 1.

4.4 Research Design

Quantitative research was conducted by analysing the datasets provided by the company. This quantitative research included the analysis of the reported emails over time and an in-depth analysis of the email content in the dataset.

4.4.1 Patterns over time. Locating the timing of the E-learning and simulation wave in the plots over time provided the opportunity to see the changes in the factors in relation to the training. Each email has a timestamp of the moment of reporting. To study the effect of the training on the reported emails, we used the timestamp to analyse the patterns. The email counts are aggregated per week or day, depending on the intended goal of the analysis. Then, the counts of the items in the group were put into a data frame and plotted over time with the counts of emails plotted by the characteristic of interest. After thorough experimentation, we opted to examine specific characteristics: classification; unique reporters; unique reports.

4.4.2 Classification. The classification of an email shows the accuracy of the employees in reporting emails. Comparing true and false positives can provide insight into high benign (No Threat Detected) report rates in the company established in Table 1, making malicious and benign classifications of particular interest. Therefore, we analysed the difference in reporting rates between emails classified as benign and malicious. This was done in relation to the two types of training the employees received, namely the E-learning and simulation wave.

4.4.3 Correlation. Besides the visual cues of seeing counts change over time, it is possible to determine the correlation between the two variables. Pearson's Correlation Coefficient provides a simple and commonly used metric to measure linear correlation using Equation 1, as below. In this research, x is the measurement of benign emails, while y denotes the measurement of malicious emails.

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{1}$$

The value lies between -1, meaning complete negative correlation, to 1, meaning complete positive correlation. 0 indicates there is no correlation. Generally, a coefficient between 0 and 0.4 indicates a weak or no correlation, between 0.4 and 0.7 a moderate correlation and higher than 0.7 a strong correlation [41].

4.4.4 Unique reporters. The number of unique people who report an email can be a metric of the security behaviour in the company. With this metric, it is possible to see whether new people report emails or if the same people keep reporting emails with few additional employees using the report function. Additionally, the number of unique reporters can increase over time, indicating a security culture where various employees participate in reporting emails.

4.4.5 Unique reports. We can gain an understanding of the underlying behaviour of employees if we compare unique reports with the total reports. However, unique reported emails are ambiguous to group as unique because emails with duplicate titles can have different content. Additionally, attackers can change email addresses and headers slightly to circumvent filters [4]. We chose to group the emails based on their title, but only if the emails were sent within two days of each other. This period is determined because most phishing campaigns only last for a short period, less than a day [15]. Additionally, 95% of the reported emails in the data are reported within a day, with only a few outliers.

4.5 Email Text Comparison

For the text analysis, emails first had to be pre-processed, after which we used several techniques to compare the text in the emails. The majority of the emails were in English (76.7 %) and evenly distributed over the data in terms of classification. Multiple steps were taken during the preprocessing, using the NLTK library in Python [7]: All non-alphanumeric characters were removed; Stopwords were removed; The words were lemmatised (to return words to their root meaning [44]).

4.5.1 Topic modelling. The chosen topic modelling model operates on numeric data instead of raw text. Therefore, every word got a unique ID to make comparison possible. Next, a LDA model [8] was used to discover the latent topics in the emails. This type of model is the most used for this type of analysis [5]. The LDA model sees the text as a mixture of all the topics. The model's outcome represented the 'topics' that best represent the information in them [31]. A downside of the LDA model is that some topics may be difficult to interpret, influencing how well the results can be interpreted. While an important limitation, the study's exploratory nature calls for flexible methods.

Analysing the topics could be done for subgroups of the data to answer part of the main research question, and to explain potential differences. The benign and malicious emails were analysed separately. The topics were then compared to see if there were striking similarities or differences in the words, indicating the most popular topics in the emails.

4.5.2 Text comparison. Besides extracting topics from the text, the emails were compared based on the similarity in their email body, to add to the interpretation of the main research question. The benign and malicious reported emails were compared to the content of the emails used in the simulation wave. The most frequently used method to compare emails on their content is TF-IDF [39]. With this method, text files are converted to a vector representation. The resulting vectors can then be compared with the cosine similarity score.

The cosine similarity score is often used to measure document similarity in text analysis [1]. First, a piece of text must be converted to a vector representation to apply cosine similarity. The score ($sim(A, B)$) is then calculated with Equation 2, where A and B are the two texts we compare, but in their vector representation. The similarity score ranges from 0, indicating no similarity between the two documents, to 1, meaning exactly the same.

$$sim(A, B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \cdot \sum_{i=1}^{n} B_i^2}} \tag{2}$$

4.5.3 Component comparison. The research question can be answered further by analysing the components in an email. Numerous components relate to E-learning, and an employee is taught to look at them to identify the nature of an email, e.g., the technical details of the email, emotions, cues, and relevance. Most components had to be extracted from the body of the email, as they were not provided separately for each email. The components in the analysis were based on previous research and preliminary documents from the company, to minimise confirmation and success bias. The preliminary documents include a difficulty rating the company uses to determine the difficulty of their simulation emails. These components allowed us to identify how many cues an employee could use to determine the email's classification. We were then able to compare the components in the simulation emails with those in the reported emails.

4.6 Ethical Considerations

The research was approved by The Human Research Ethics Committee (HREC) of the research institution. The E-learning data is not linked to individuals, and no analysis was done on individual behaviour by linking reports to personal information. Due to the confidential nature of the data, not all results will be publicly accessible. Measures were taken to guarantee confidentiality along with proper results. To safeguard the identity of the employees who reported the emails, the results were aggregated and cannot be linked back to individuals.

Strict data protection protocols were followed throughout the research process to ensure the secure handling of sensitive data. Discussions and collaboration regarding the data and code were limited to the internal research team at the cooperating company and authorised individuals. Additionally, an expert from the company considered the choices made in the data-cleaning process. Furthermore, any external discussions involving the data were conducted under the appropriate confidentiality agreements and in compliance with the company's data protection policies. These measures were taken to safeguard the organisation's and its employees' privacy and security.

5 Results

Here we detail the results of our analysis of the dataset of reported emails, focusing on the progression of reports over time for distinct categories of reports.

5.1 E-Learning and Phishing Simulations

We plotted the number of E-learning courses completed per week in Figure 1. The dataset contains all emails employees have reported between January 1st, 2022, and April 1st, 2023. The first E-learning was completed on February 3, 2022. Employees were given four weeks to complete the training, and most employees completed the training in the mandatory period; this is represented as the highlighted spike and diminishing peak in the figure, as all employees gradually complete the training. However, some employees complete the training later in the year due to external factors (such as e.g., a sabbatical or pregnancy, resulting in the employee's absence). After the introduction of the E-learning, new employees were required to also complete the training as part of their onboarding.

Besides the mandatory E-learning, employees received simulated phishing emails corresponding to email components discussed in the training. The company-wide phishing simulation campaign started on November 14 2022, and ended on November 25 2022. A smaller campaign was performed at the end of February 2022, and limited to one division. In April 2023, there was a simulation in one of the Tech hubs of the company. Additionally, throughout the year, there are several simulations sent to employees (limited to specific countries). An overview of all reported simulation emails and the duration of the simulation wave can be found in Figure 2. With this information, we can examine the development of the reported emails over time in relation to the training the employees have received.

5.2 Progression of Phishing Reports over Time

The analysis of the data has been split into two parts. First, the effect of the simulation wave on the reports over time is examined, to get an understanding of employee behaviour and the relation with the training.

5.2.1 Classification. The peak in the simulation emails, as shown in Figure 2, is highlighted over the period November 14 to November 25 of 2022. Figure 3 shows the number of reported emails within each classification from Table 1, over the dataset time period. In Figure 3, the simulation emails have been removed from the data.

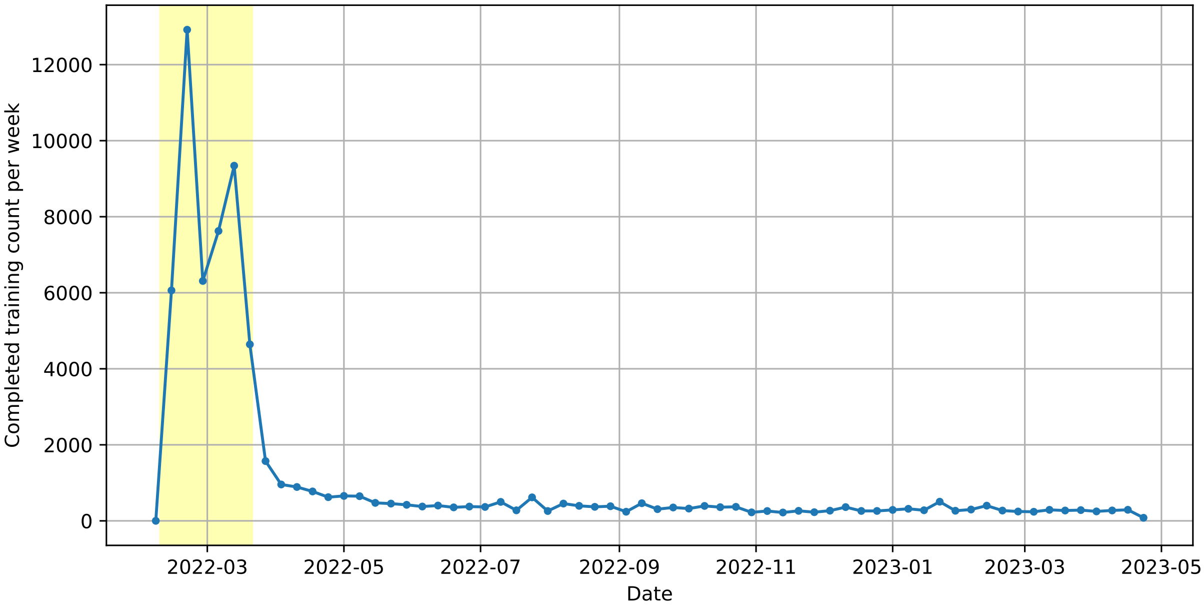

Figure 1: The progression of E-learning completion over time, between February 3rd, 2022 and April 1st, 2023. The yellow highlight cover an approximate month-long period starting February 3rd 2022, as E-learning was introduced. The counts are accumulated per week, where each dot represents the Sunday of the week the training was completed.

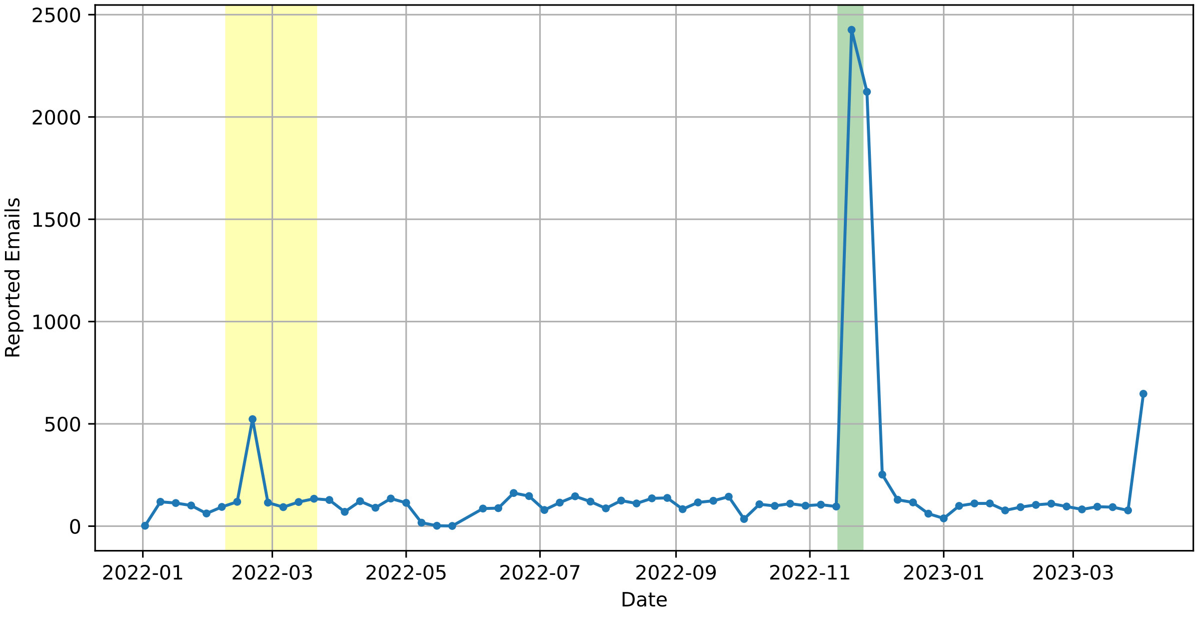

Figure 2: The progression of reported simulation emails over time, between January 1st 2022 and April 1st 2023. The yellow highlight cover an approximate month-long period starting February 3rd 2022, as E-learning was introduced. The green highlight over the period November 14, 2022 to November 25, 2022 represents the wide-scale wave of phishing simulation emails.

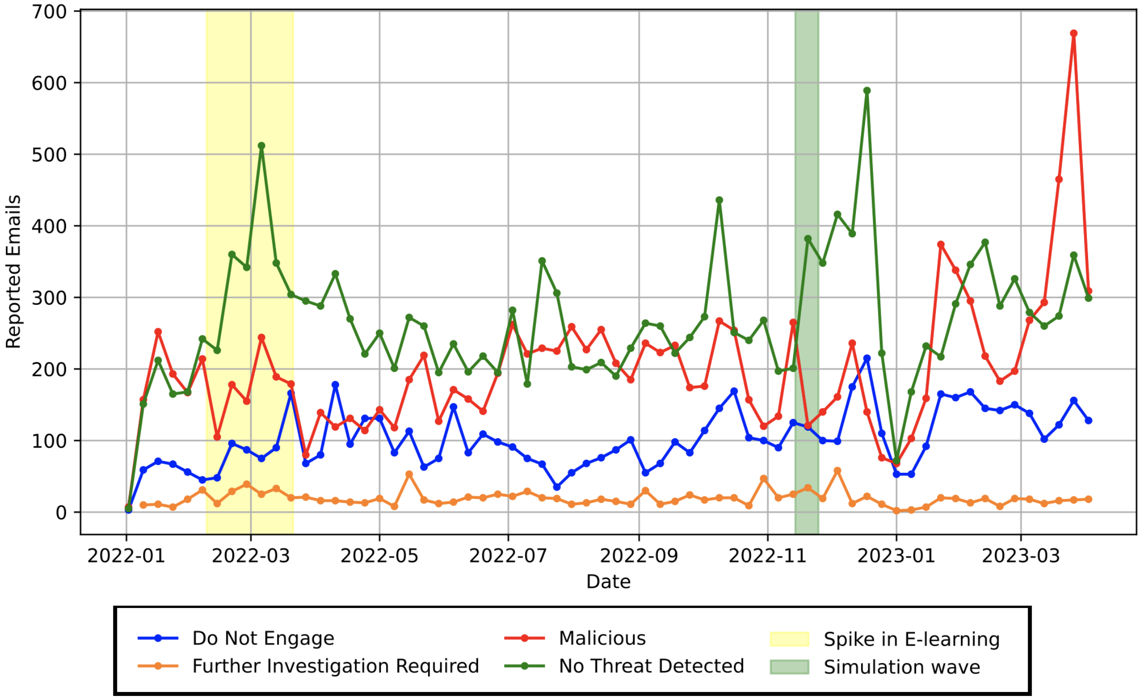

Figure 3: The progression of reported emails, where the simulation emails are not included in the presented data. The counts are accumulated per week, where each dot presents the Monday of the week. The legend shows the classification of the emails and the two time periods for the training and simulation wave highlighted.

Most notable is that the reported emails are mainly classified as No Threat Detected or Malicious, as already seen in Table 1.

There are several other notable points in Figure 3: for the benign emails, this includes a period during training and after the simulation wave, as a potential 'after effect' of heightening attention to the possibility of emails being malicious. It has been noted elsewhere that increased reporting can result in a spike in helpdesk requests [10]. There is also a drop in reports at the end of December, explained by the many employees on holiday during this period.

The reporting trends outside of these notable events may appear not to follow any particular trend, but attacks are not subject to regular patterns, as also seen in phishing emails in, e.g., a university setting [36]. It is clear that the rate of reporting for No Threat Detected is comparable to the number of Malicious emails reported over time, meaning that the company is tolerating a notable rate of false-positive reports (emails being reported that in a sense did not need to be reported). This is mirrored in the regular reporting of suspicious emails in the Do Not Engage category, where an email may not have had malicious content, but employees reported the email regardless. This is reflected in the overall classifications of reported emails over the observed time period, as in Table 1; No Threat Detected was 34.8%, and the combination of Malicious, Do Not Engage, and Further Investigation Required emails was 41.9%.

Notably for efforts to involve users as 'human as a sensor' or 'crowd-sourced' defense (as for phishing, e.g., [11]), to directly contribute to an organization's threat detection, there were a regular – but comparatively very small – number of emails reported that were classed as Further Investigation Required where automatic tools did not immediately classify the email once reported, i.e., the user identified a potential threat that the company's dedicated security apparatus did not. This is notable given the comparatively high number of No Threat Detected emails reported alongside those emails.

We can zoom in on the area around the simulation weeks in Figure 4. The days of the week can be seen clearly, as the reports draw near zero during the weekend. There is no increase in Malicious reported emails after the simulation compared to before. The benign emails do seem to increase during and after the simulation.

5.2.2 External factors. External factors can influence reporting activity. The events over time that can be analysed can be found in Figure 5. Throughout the year, newsletters are sent to employees. Often they contain topics related to phishing emails to keep employees engaged. Of note also is the peak in benign reports in mid-December, as visible in both Figure 3 and Figure 4. This peak can be attributed to a specific benign email reported by multiple employees. This specific email had asked employees to change their expired passwords and giving them 60 minutes to do so.

5.2.3 Pearson's correlation coefficient. Pearson's correlation coefficient between benign (No Threat Detected) and malicious reported emails with the count per week is 0.247. This indicates no or a weak correlation. The same factor with the data aggregated per day is 0.465, indicating a moderate correlation between the reports of benign and malicious emails.

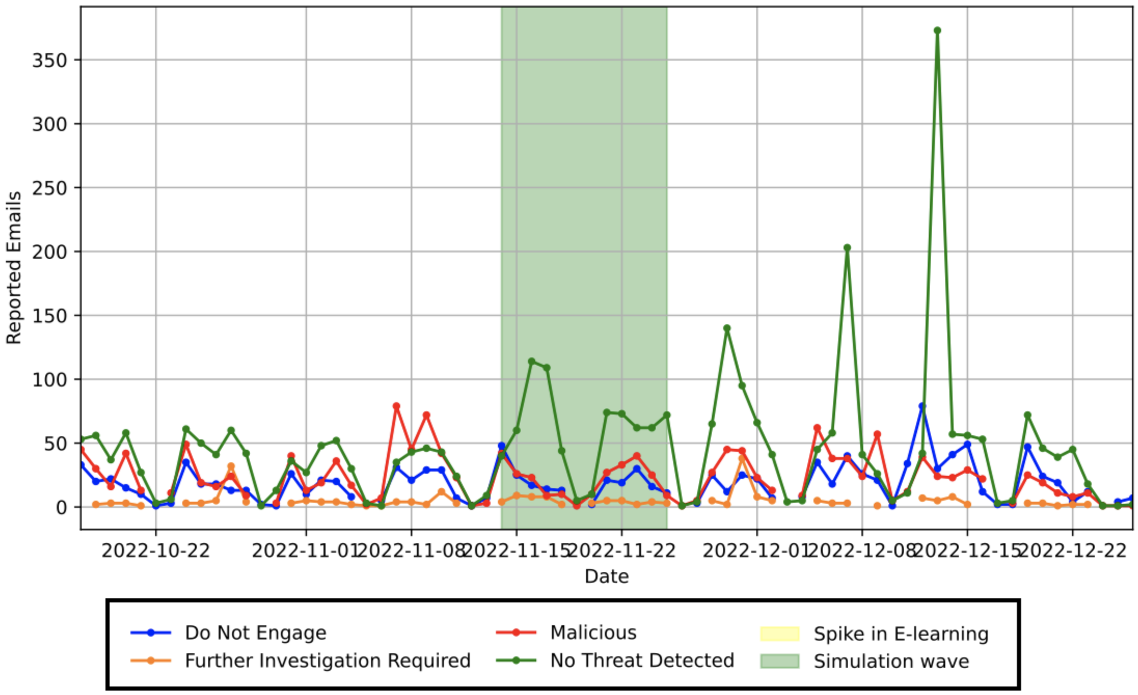

Figure 4: The progression of the reported emails, where the simulation emails are not included in the presented data. The counts are portrayed per day. There has been zoomed in to the four weeks before and after the simulation.

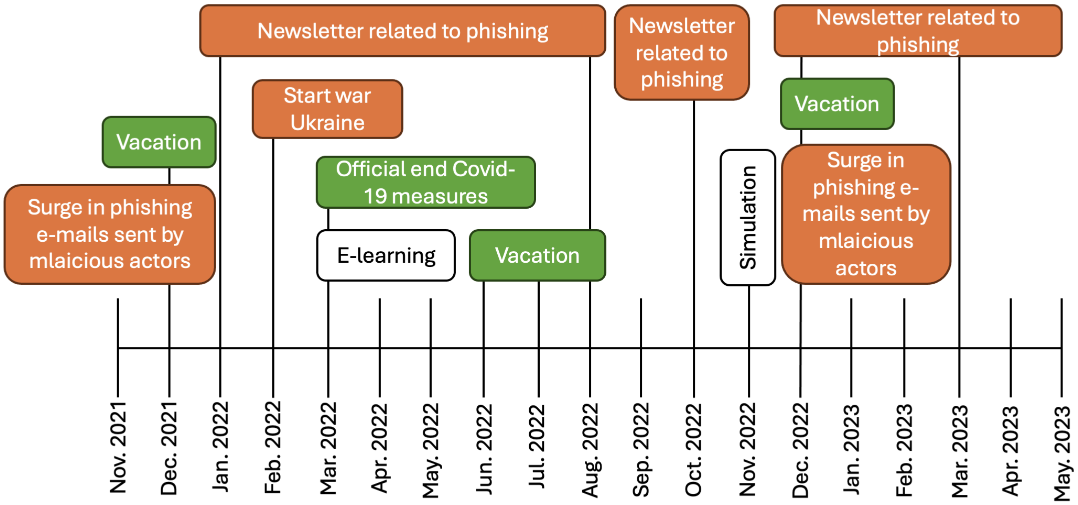

Figure 5: A timeline of events or circumstances that can influence the reporting behaviour of employees. An orange box indicates an expected higher reporting rate. A green box indicates an expected lower reporting rate. The E-learning and the simulation wave are represented in white boxes as these are under consideration during the research.

5.3 Unique Reported Emails

To understand the progression of reported emails, metrics besides total number of reported emails can provide insight.

We use the number of *unique reported emails* over time to explain outliers present in the data. The unique reported emails per week show that the number of unique emails fluctuates less than the total number of reported emails in Figure 6. We classify an email as unique if it has not been reported in two days of another email with the same title.

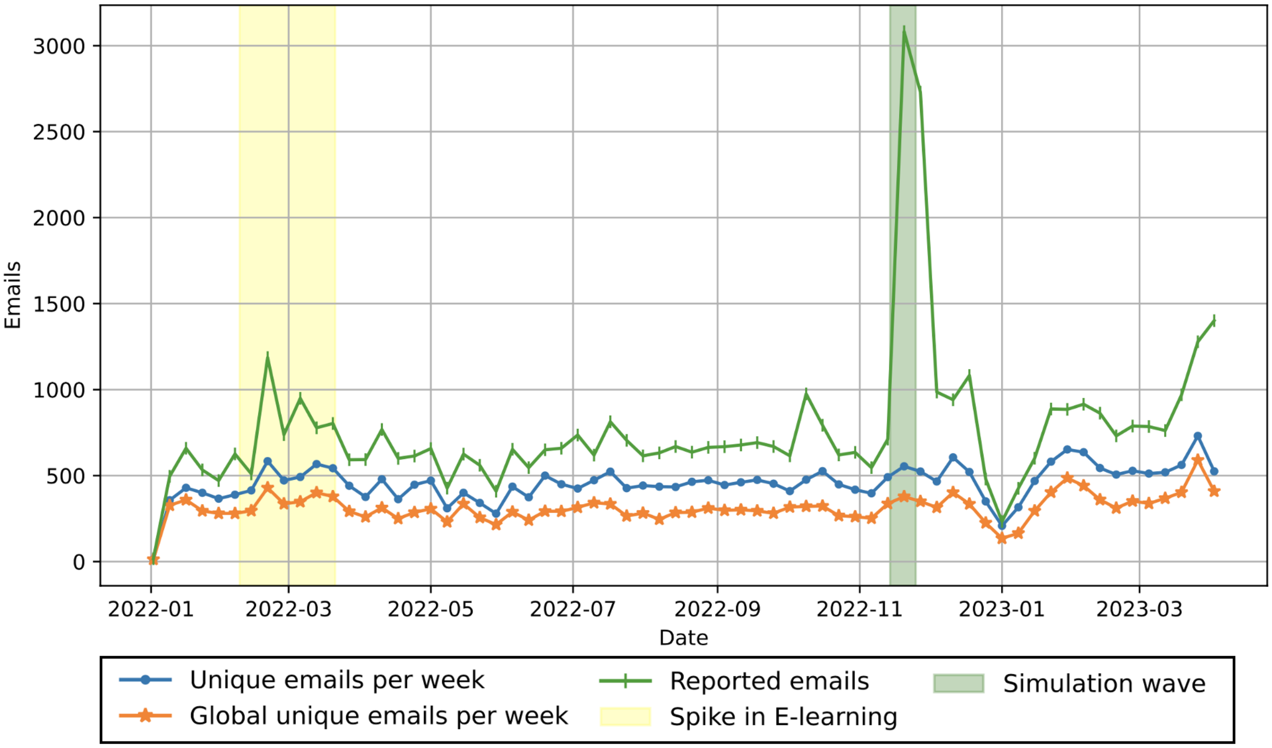

Figure 6: The unique reported emails over time. Additionally, the total email count and the global unique emails are shown. An email is globally unique if the title has not been reported in the data before. An email is unique if it has not been reported within two days of another email with the same title.

Alternatively, the *global unique emails* show the unique titles throughout the observed period. These globally unique emails show the same pattern. The spikes in the reported emails can be explained by one or some specific emails that more employees have reported. This situation can occur only if the emails are sent to numerous employees, which could happen in the case of a harmless mass email or a phishing attack aimed at multiple employees (indicating that a measurement of unique report can also capture some details about the kinds of phishing campaigns being launched against an organization).

From these unique emails, the classification is known, as the emails with the same title have been given the same classification throughout the dataset. Looking at the unique emails per classification in Figure 7, the changes in reported emails are distinctly different than those seen in Figure 3. There is a peak in unique benign reported emails during the E-learning. The training seems to correlate with an increase in benign reported emails. The unique reported malicious emails peak in July 2022 and February and March 2023.

5.4 Unique Reporters

Understanding the reports over time continues with understanding the employees who report emails. Every week, around 500 unique employees report an email, as seen in Figure 8. During the November 2022 simulation wave (but also the March 2023 wave that is not highlighted), there is an apparent increase in unique reporters. The changes in the number of unique employees who reported an email follow a similar pattern as the changes in reports, more closely than unique emails. The increase in reports can be explained by multiple employees reporting the same email(s). During E-learning, there is also a peak in unique reporters, though we also see that these spikes diminish. Lain et al. [35] also noted a constant rate of reporting; in our partner organization users received an immediate confirmation, which may also tally with [35], who found that users who received confirmation of a true positive continued to report.

5.5 Component Comparison

Another comparison can be made on the components present in the emails. The comparison between the simulation, malicious and benign emails is provided in Table 2. For security reasons, the components are not specified to great detail, though we relate the components to, for instance, ‘email presentation’ as a broad category of factors identified by Zhuo et al. [45]. ‘Simulation all’ includes all of the simulation campaigns, and ‘Simulation wave’ refers to the wide-scale simulation campaign of November 2022. ‘Focus in E-learning’ refers to whether employees were informed about the Component during training. All component classifications are informed by prior research on technical aspects of an email [33], emotions [9, 43, 45], cues [42, 43], and relevant or expected emails [29]. Components 10 to 12 look into the presence of a topic

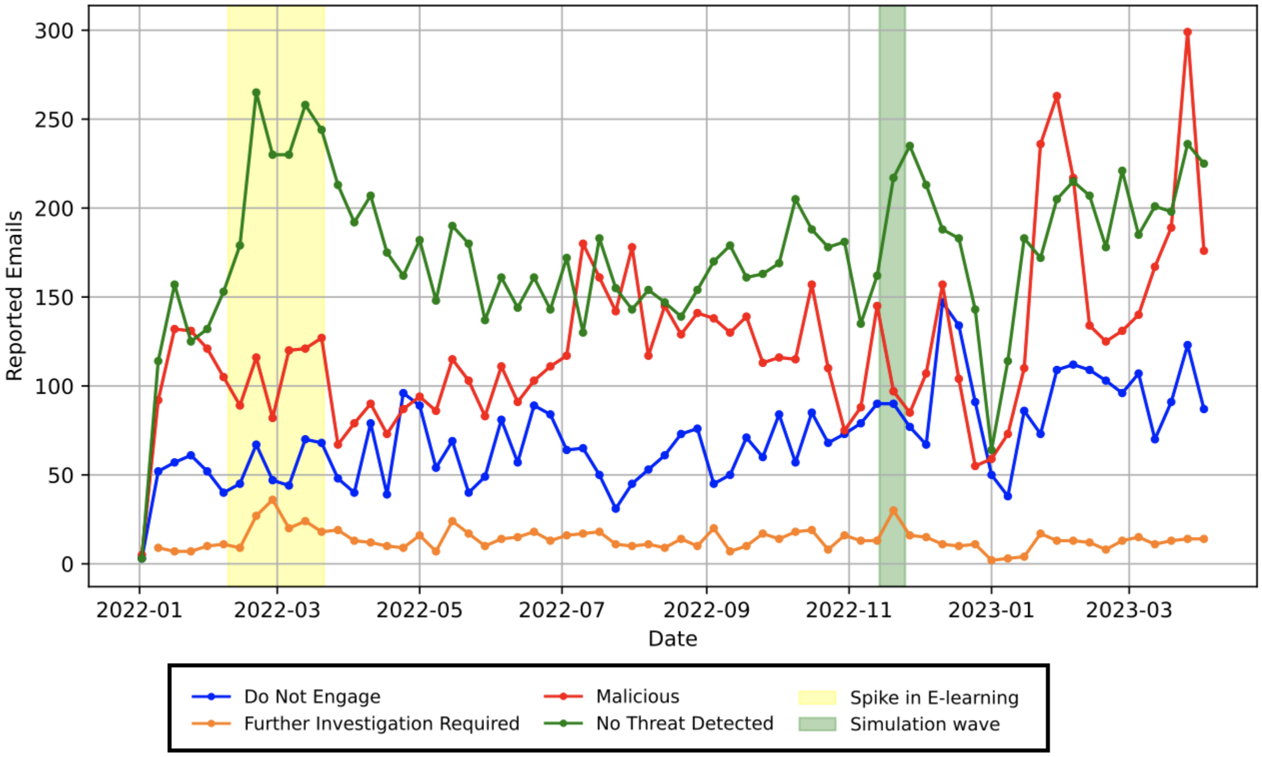

Figure 7: The unique reported emails per classification.

in the emails. As there are only ten emails in the simulation wave, the percentages of this column are multiples of 10.

Several components of Table 2 are present in all simulation emails during the wave in November, while the reported emails contain lower percentages of these components. Component 12 refers to words associated with a topic. Benign emails frequently mention this topic, while malicious emails have a lower frequency of these words.

The results indicate that there can be overlap in Component presence in both benign and malicious emails – only in one instance (Component 3) were malicious emails distinguished more from benign emails (and even there, marginally so). This also means that there is a high overlap between the content of reported benign and malicious emails, such that it would be difficult to hold on to any hope of reducing reporting to precisely malicious emails (and managing to not report any benign emails). What is also of note is that many topics were featured to a lesser degree in the simulations, and were not a topic of focus in the E-learning, yet featured in a number of benign email reports, such as Components 4, 8-9, and 11-13 – employee reports also caught malicious emails with these components, potentially hinting at how employees were not simply 'pattern matching' the emails featured in training, but adapting and applying what they were learning (which is as far as we are aware, has not been researched to any great degree up to now).

6 Discussion

6.1 Reporting is about Suspicious Emails, not Only Malicious Emails

From the descriptive statistics in subsection 4.3, the percentage of 34.8% of emails classified as *No Threat Detected* compared to a combined 39.4% for *Malicious* and *Do Not Engage* is striking. Employees report almost as many benign emails as 'correctly' reported emails. It is not trivial to clearly distinguish benign from malicious emails, and many benign emails contain suspicious attributes similar to malicious emails. This comparison articulates in practice what it would mean for the analysis of reported emails, considering the extent to which employees are encouraged to 'err on the side of caution' in reporting (e.g., when they are not sure if an email is malicious or not).

Several benign emails have many components that an employee may attribute to a phishing email. In Table 2, it can be seen that the reported benign emails contain components highlighted in E-learning as email elements that could be suspicious. Employees receive feedback on the emails they reported and their true classification (malicious or not), indicating if an email was reported correctly. Given that there can be such high rates of false-positives, the nature of feedback may need to move beyond true or false phishing, to reaffirming employees that they should still report emails that have suspicious components. Otherwise, an employee who reports benign emails may interpret their actions as incorrect, and report less, where Lain et al. [35] noted such an effect. Reporting

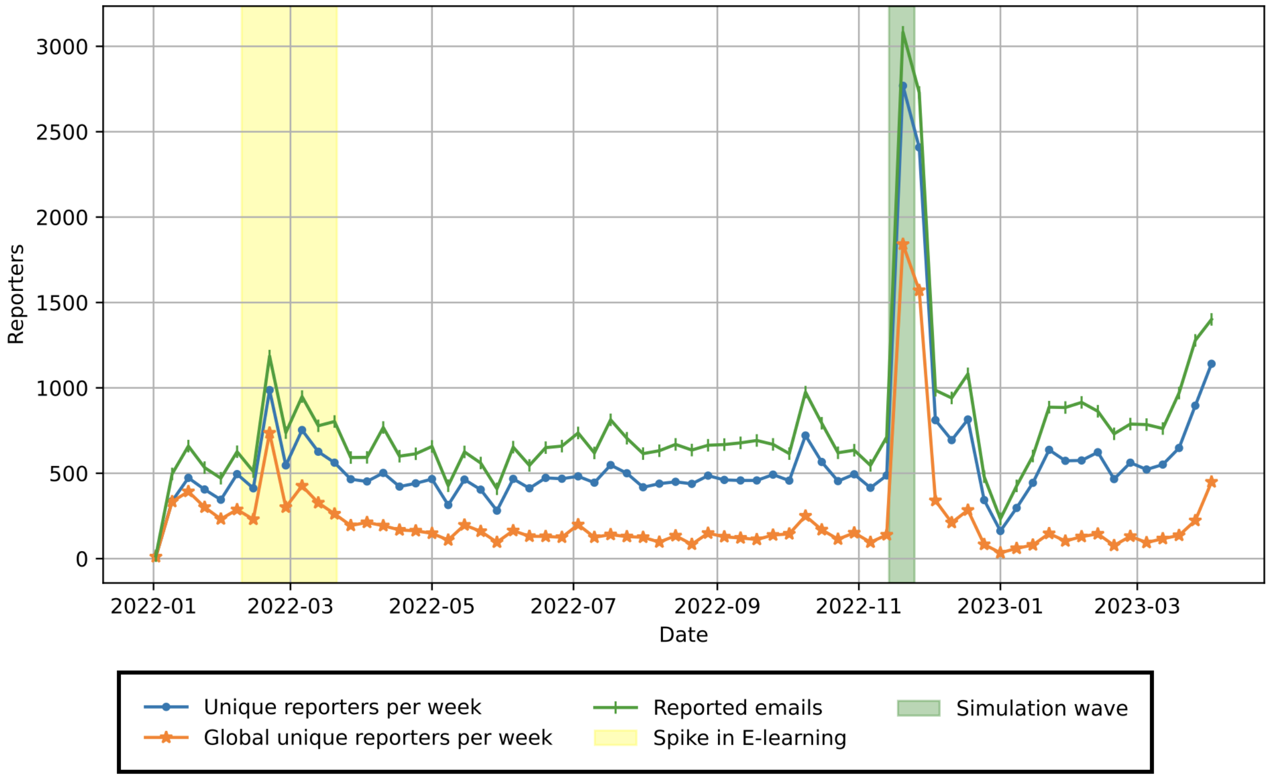

Figure 8: The unique reporters aggregated per week. Blue shows the unique reporters per week, whereas orange shows the addition of new reporters from the beginning per week.

Table 2: Percentage of emails where the component is present in the body of the text. The last column specifies whether specific attention is spent on this component during the E-learning.

Component	Benign	Malicious	Simulation all	Simulation wave	Focus in E-learning
1. English language	81 %	59 %	100 %	100 %	No
2. Technique: Link manipulation	93 %	65 %	100 %	100 %	Yes
3. Technique: company specific link	59 %	69 %	72 %	100 %	Yes
4. Personalised email	39 %	11 %	51 %	30 %	No
5. External sender	75 %	86 %	76 %	100 %	Yes
6. Technique: sender manipulation	30 %	18 %	54 %	60 %	Yes
7. Technique: Time pressure	80 %	72 %	68 %	70 %	Yes
8. Technique: Whaling	52 %	29 %	37 %	30 %	No
9. Contains company terminology	60 %	40 %	32 %	30 %	No
10. Topic: Security	11 %	10 %	13 %	20 %	Yes
11. Topic: Pay	14 %	10 %	9 %	20 %	No
12. Topic: HR	60 %	32 %	55 %	60 %	No
13. Topic: Known service	19 %	19 %	37 %	50 %	No

over time is then arguably not about becoming more accurate at *exactly* spotting phishing, if malicious and benign emails will always overlap. It is then about the organization supporting employees to report benign emails and *ensuring* quick confirmation, so that they can return to work that potentially relies on benign emails that they were unsure of.

This finding adds to the research of Steves et al. [40] and the NIST Phish Scale, where *premise alignment* could be measured over time for the organization to determine how much benign reporting it can accept. There is then a certain level of ‘incentivized over-reporting’, where the reporting cost is not per malicious email, but per *suspicious* email (i.e., malicious emails and benign emails with ‘suspicious-looking’ components).

6.2 Unique Reporters

The relation between the number of unique reporters and the total reported emails in Figure 8 provides insight into the company's security behaviour [22]. During the three spikes in simulation waves, in February 2022, November 2022 and March 2023, as seen in Figure 2, there are noticeable peaks in the number of unique reporters. This shows that the simulation emails manage to reach employees who do not report other emails during the year, as seen by the peaks in the orange line in Figure 8.

The development over time of the unique reporters also shows whether employees who report are familiar with the process of reporting emails before the simulation. This is a new perspective towards the way simulation waves can be used. Where prior research [35] looks at the reporting rates separately, combining these rates with the actual reports shows a broader picture. We find that 'reporter uniqueness' emerges as a visible measure of security-related culture after training has been deployed. Such a measure has up to now had little consideration in practice. Aiming always to increase the number of people reporting a particular phishing campaign or email has an unclear end-goal. Alternatively, knowing that there is a *reasonable* number of different users reporting each email is more tractable, as it arguably only takes one person to report an email – we saw a few hundred new reporters every week, as in Figure 8.

6.3 User-centred Reporting Measures

The total reported emails per classification in Figure 3 show no attributable change in the reporting behaviour of employees before and after the phishing simulation. However, looking at the Component overlap in Table 2, there is a need to consider a kind of 'reporting sensitivity' or amplification of reporting, as in Figure 3 and Figure 4. If high reporting sensitivity were to be encouraged in an organization, the time between reporting and getting a confirmation about the reported email (if at all) becomes critical – a longer wait could have consequences for the working of the organization. Legitimate (benign) emails would presumably be avoided until they are checked and confirmed to be harmless. This would constitute an unintended harm of the reporting process [14], relying on there being a quick report-and-confirmation process, rather than relying on only employee engagement and effort.

6.4 Component Comparison

The connection of the content comparison with the E-learning shows that components not discussed in the E-learning are generally less represented in the reported malicious emails, and the focus of the simulation emails seems to lie with the factors discussed in the E-learning. Several components in the reported emails are underrepresented in the simulation emails. Relating these components to the metrics discussed in the E-learning links the two together. As stated, some of the reported malicious emails differ from the E-learning, as they contain additional or missed components that are present in all simulation emails but are also not discussed during the E-learning. This could indicate that employees can extend their knowledge to other types of emails or, prior to the learning, already knew of email components that could warrant reporting an email.

6.5 Limitations

Our data was limited in terms of what behaviours we could analyze before the widespread E-learning early in the dataset. For instance, E-learning may have had an effect and simulations act to 'top up' reporting skills. We were, however, able to observe effects from simulation waves, and after-effects of both E-learning and simulation waves.

To determine unique emails over time, the titles of the emails have been cleaned and compared to group the identical titles. While this tactic is quite effective for benign emails, as campaigns are the same for each employee, the approach is likely less suitable for detecting malicious campaigns which actively try to avoid easy grouping [32].

Future work will separate reports according to departments; the complex department associations analyzed here did not lend themselves readily to linking modes of working and asset access to a discernible business line.

6.6 Recommendations

Here we provide initial recommendations for further research, based on our findings.

Large-scale analysis of anti-phishing reporting behaviour and adapting simulations. Many large-scale analyses identify prevalent components of phishing emails; many smaller-scale analyses involve humans in studies of their responses to suspected-phishing emails; few combine the two, to analyze reporting data from a human-centred perspective. Although some emails might not be reported [16], comparing reported emails over time can inform the design of simulation emails.

Honour training with reporting-outcome confirmation. Much related research focuses on emotional and behavioural cues, and not the need for simplicity in reporting and the need for feedback. Costs are often discussed in terms of the costs of the training and simulations [35]. We saw a high rate of benign reports alongside valuable reports of malicious emails; reporting of benign emails went up temporarily after training and simulation events. The rapid confirmation provided to employees (Section 3) potentially saved work time from being lost to waiting on whether the email was genuine (and could be acted on), where a fast response can also boost reporting [43].

Surface the non-security costs of cautious reporting. Security-minded practice would naturally encourage users to err on the side of caution when considering whether to report. This means that there will always be benign reports. If phishing reporting is inherently cautious, then there are costs incurred from the disruption that reporting causes to benign emails, e.g., when an email requesting reset of account credentials is genuine.

7 Conclusions

Here we reported on a collaboration with a partner bank, to examine a repository of many thousands of reported emails from a behavioural perspective. We found that the level of reporting of benign emails is comparable to the number of malicious emails reported. We also saw limited evidence of training and simulations 'amplifying' the reporting of benign emails. Several hundred benign and known-malicious emails may be reported on a week-by-week basis, capturing several emails which require further investigation. We also uncovered metrics for unique reporters per email campaign. Future work has the potential to focus in on the differing reporting patterns of specific departments; such work could differentiate between teams who have access to various kinds of important data and systems, or who regularly interact with external entities with whom they have limited prior history. These are arguably the users who have the most pressure and need the most support to report with an abundance of caution, while also minimizing any disruption to their work from reporting.

Acknowledgments

We wish to thank the partner organization for their cooperation. We would like to thank Alessandro Giordani for his help with the data analysis.

References

[1] Shivam Agarwal. 2013. Data mining: Data mining concepts and techniques. In *2013 international conference on machine intelligence and research advancement*. IEEE, 203–207.

[2] Adel Ismail Al-Alawi and Sara Abdulrahman Al-Bassam. 2019. Assessing The Factors of Cybersecurity Awareness in the Banking Sector. *Arab Gulf Journal of Scientific Research* 37, 4 (2019), 17–32.

[3] Issam Al-Shanfari, Warusia Yassin, Raihana Syahiraha Abdullah, Nabil Hussein Al-Fahim, and Roesnita Ismail. 2021. Introducing a novel integrated model for the adoption of information security awareness through control, prediction, motivation, and deterrence factors: A pilot study. *Journal of Theoretical & Applied Information Technology (JATIT)* (2021).

[4] Joseph Aneke, Carmelo Ardito, and Giuseppe Desolda. 2019. Designing an intelligent user interface for preventing phishing attacks. In *IFIP Conference on Human-Computer Interaction*. Springer, 97–106.

[5] Claus Boye Asmussen and Charles Møller. 2019. Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data* 6, 1 (2019), 1–18.

[6] A. Bhardwaj, V. Sapra, A. Kumar, N. Kumar, and S. Arthi. 2020. Why is phishing still successful? *Computer Fraud & Security* 9 (2020), 15–19. https://doi.org/10.1016/S1361-3723(20)30098-1

[7] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc. https://www.nltk.org/book/

[8] David M Blei. 2012. Probabilistic topic models. *Commun. ACM* 55, 4 (2012), 77–84.

[9] Mark Blythe, Helen Petrie, and John A Clark. 2011. F for fake: four studies on how we fall for phish. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3469–3478.

[10] Lina Brunken, Annalina Buckmann, Jonas Hielscher, and M Angela Sasse. 2023. "To Do This Properly, You Need More Resources": The Hidden Costs of Introducing Simulated Phishing Campaigns. In *32nd USENIX Security Symposium (USENIX Security 23)*. 4105–4122.

[11] Pavlo Burda, Luca Allodi, and Nicola Zannone. 2020. Don't forget the human: a crowdsourced approach to automate response and containment against spear phishing attacks. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 471–476.

[12] Deanna D. Caputo, Shari Lawrence Pfleeger, Jesse D. Freeman, and M. Eric Johnson. 2014. Going Spear Phishing: Exploring Embedded Training and Awareness. *IEEE Security & Privacy* 12, 1 (2014), 28–38. https://doi.org/10.1109/MSP.2013.106

[13] Chris Chatfield. 1995. *Problem solving: a statistician's guide*. CRC Press.

[14] Yi Ting Chua, Simon Parkin, Matthew Edwards, Daniela Oliveira, Stefan Schiffner, Gareth Tyson, and Alice Hutchings. 2019. Identifying unintended harms of cybersecurity countermeasures. In *2019 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 1–15.

[15] Catalin Cimpanu. 2020. Phishing campaigns, from first to last victim, take 21h on average. ZDNet. https://www.zdnet.com/article/phishing-campaigns-from-first-to-last-victim-take-21h-on-average/

[16] Arnout Van de Meulebroucke. 2021. https://phished.io/es/blog/an-end-to-pride-and-prejudice-everyone-is-susceptible-to-phishing

[17] De Nederlandsche Bank N.V. 2022. Jaarverslag 2022, koers houden. https://www.dnb.nl/publicaties/publicaties-dnb/jaarverslag/jaarverslag-2022/

[18] European Banking Authority. 2018. Guidelines on security measures for operational and security risks under the PSD2. https://www.eba.europa.eu/regulation-and-policy/payment-services-and-electronic-money/guidelines-on-security-measures-for-operational-and-security-risks-under-the-psd2

[19] Aoife Feeley, Matthew Lee, Michelle Crowley, Iain Feeley, Ryan Roopnarinesingh, Sinead Geraghty, Brian Cosgrave, Eoin Sheehan, and Khalid Merghani. 2022. Under viral attack: An orthopaedic response to challenges faced by regional referral centres during a national cyber-attack. *The Surgeon* 20, 5 (2022), 334–338.

[20] Anna Georgiadou, Ariadni Michalitsi-Psarrou, and Dimitris Askounis. 2022. Cyber-Security Culture Assessment in Academia: A COVID-19 Study: Applying a Cyber-Security Culture Framework to assess the Academia's resilience and readiness. *ACM International Conference Proceeding Series* (8 2022). https://doi.org/10.1145/3538969.3544467

[21] Sanjay Goel, Kevin Williams, and Ersin Dincelli. 2017. Got Phished? Internet Security and Human Vulnerability. *Journal of the Association for Information Systems* 18 (1 2017), 2. Issue 1. https://doi.org/10.17705/1jais.00447

[22] Kristen Greene, Michelle Steves, and Mary Theofanos. 2018. No Phishing beyond This Point. *Computer* 51 (6 2018), 86–89. Issue 6.

[23] Jonas Hielscher, Uta Menges, Simon Parkin, Annette Kluge, and M Angela Sasse. 2023. "Employees Who Don't Accept the Time Security Takes Are Not Aware Enough": The CISO View of Human-Centred Security. In *32nd USENIX Security Symposium (USENIX Security 23)*. 2311–2328.

[24] Doron Hillman, Yaniv Harel, and Eran Toch. 2023. Evaluating Organiza-tional Phishing Awareness Training on an Enterprise Scale. *Computers & Security* (2023). https://doi.org/10.1016/j.cose.2023.103364

[25] ING group N.V. 2022. Annual Report 2022. https://www.ing.com/Investor-relations/Financial-performance/Annual-reports.htm

[26] Daniel Jampen, Gürkan Gür, Thomas Sutter, and Bernhard Tellenbach. 2020. Don't click: towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences* 10 (12 2020). Issue 1.

[27] S. Kaddoura, G. Chandrasekaran, D. Elena Popescu, and J. H. Duraisamy. 2022. A systematic literature review on spam content detection and classification. *PeerJ. Computer science* 8 (2022), e830. https://doi.org/10.7717/peerj-cs.830

[28] Hwee-Joo Kam, Thomas Mattson, and Sanjay Goel. 2020. A cross industry study of institutional pressures on organizational effort to raise information security awareness. *Information Systems Frontiers* 22, 5 (2020), 1241–1264.

[29] Abu Kamruzzaman, Kutub Thakur, Sadia Ismat, Md Liakat Ali, Kevin Huang, and Hasnain Nizam Thakur. 2023. Social Engineering Incidents and Preventions. (2023). https://doi.org/10.1109/CCWC57344.2023.10099202

[30] Kamlesh Kanwal, Wenming Shi, Christos Kontovas, Zaili Yang, and Chia-Hsun Chang. 2022. Maritime cybersecurity: are onboard systems ready? *Maritime Policy & Management* (2022), 1–19.

[31] Shashank Kapadia. 2019. Topic Modeling in Python: Latent Dirichlet Allocation (LDA). https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0

[32] Liyiming Ke, Bo Li, and Yevgeniy Vorobeychik. 2016. Behavioral Experiments in Email Filter Evasion. *Proceedings of the AAAI Conference on Artificial Intelligence* 30, 1 (2 2016). https://doi.org/10.1609/aaai.v30i1.10061

[33] Danielle Kelvas. 2023. SLAM Method: How to Prevent HIPAA Email Phishing Attacks. https://www.hipaaexams.com/blog/slam-method#:~:text=The%20SLAM%20method%20is%20an,Link%2C%20Attachment%2C%20and%20Message.

[34] Artur Kuchumov, Elena Pecheritsa, Alexandra Chaikovskaya, and Elena Maslova. 2022. Digitalization of Economics: Modern Financial Technologies and Their Influence on Economic Security. In *IV International Scientific and Practical Conference (DEFIN-2021)*. Association for Computing Machinery, Article 28, 7 pages. https://doi.org/10.1145/3487757.3490866

[35] Daniele Lain, Kari Kostiainen, and Srdjan Capkun. 2022. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. *Proceedings - IEEE Symposium on Security and Privacy* 2022-May (2022), 842–859.

[36] Ethan Morrow. 2024. Scamming Higher Ed: An Analysis of Phishing Content and Trends. *Computers in Human Behavior* (2024), 108274.

[37] K. Parsons, A. Mccormac, M. Pattinson, M. Butavicius, and C. Jerram. 2014. Using actions and intentions to evaluate categorical responses to phishing and genuine emails. *Proceedings of the 8th International Symposium on Human Aspects of Information Security and Assurance, HAISA 2014* (2014), 30–41.

[38] Andrew Reeves, Kathryn Parsons, and Dragana Calic. 2020. Whose risk is it anyway: How do risk perception and organisational commitment affect employee information security awareness? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12210 LNCS (2020), 232–249. https://doi.org/10.1007/978-3-030-50309-3_16

[39] Said Salloum, Tarek Gaber, Sunil Vadera, and Khaled Shaalan. 2022. A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. *IEEE Access* 10 (06 2022). https://doi.org/10.1109/ACCESS.2022.3183083

[40] Michelle Steves, Kristen Greene, and Mary Theofanos. 2020. Categorizing human phishing difficulty: a Phish Scale. *Journal of Cybersecurity* 6, 1 (2020), tyaa009.

[41] Sarah Thomas. 2023. Understanding the Pearson Correlation Coefficient. Outlier. https://articles.outlier.org/pearson-correlation-coefficient

[42] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao. 2011. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems* 51, 3 (2011), 576–586. https://doi.org/10.1016/j.dss.2011.03.002

[43] Emma J. Williams, Joanne Hinds, and Adam N. Joinson. 2018. Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies* 120 (2018), 1–13. https://doi.org/10.1016/j.ijhcs.2018.06.004

[44] R. Yash. 2023. Python | Lemmatization with NLTK. https://www.geeksforgeeks.org/python-lemmatization-with-nltk/

[45] Sijie Zhuo, Robert Biddle, Yun Sing Koh, Danielle Lottridge, and Giovanni Russello. 2022. SoK: Human-Centered Phishing Susceptibility. *ACM Transactions on Privacy and Security* (3 2022). https://doi.org/10.1145/3575797

Appendix: data contents

Reported phishing emails

- IncidentType: Is the same for every datapoint, namely SEA (Suspicious Email Analysis)
- IncidentStatus: The status of the incident, for all emails this is 'Closed'
- IncidentID: ID of the incident, if an email has multiple indicators these IDs are the same for the different rows. So one reported email has one IncidentID, but can have multiple data points.
- Classification: The type of email, and thus the result of the analysis. Can be: Malicious, Simulation, No Threat Detected, Do Not Engage.
- ThreatType: The type of threat that is identified in the email. Can be: Link, Response, or Payload.
- subClassification: Specification of the classification
- EmailReportedBy: The email address of the person who reported an email as phishing
- Subject: The subject header of the email
- Sender: The email address of the person who send the supposed phishing email
- Reported: Date and time the email was reported
- Modified: Not sure what this entails. Format: Date and time.
- Resolved: Time at which the email was classified and closed by the company
- Age: Time between the reported and resolved moment. Format: time in seconds
- Message ID: ID given to the message. If an email has multiple indicators these IDs are the same for the different rows.
- Indicator Type: Type of indicator that helps to indicate the email is phishing [nan, URL, email address, payload]
- Indicator: In case an indicator is present, the indicator is given here.
- Campaign ID: If the email is marked as part of a campaign the id of the campaign is given here

Department specification

- Email
- Given name
- Country code
- Organisation
- Company
- Department
- pwdLastSet: date the password has last been reset
- Manager: Specifications of the manager of the employee
- DistinguishedName

Employee training: Mandatory E-learning

- Enroll time: time the employee was enrolled for the training and thus when someone could get started
- Start time: the time an employee started with the training
- End time: The time an employee finished the training
- Status: status of the training, whether someone finished the training
- Name/email

Counterfeit phishing emails

This analysis has already been made by the company

- Email body
- Title
- Click rates
- Compromise rates
- Reporting rates