

The Impact of Workload on Phishing Susceptibility: An Experiment

Sijie Zhuo

University of Auckland
szhu842@aucklanduni.ac.nz

Robert Biddle

University of Auckland
& Carleton University, Ottawa
robert.biddle@auckland.ac.nz

Lucas Betts

University of Auckland
lbet419@aucklanduni.ac.nz

Nalin Asanka Gamagedara Arachchilage

University of Auckland
naln.arachchilage@auckland.ac.nz

Yun Sing Koh

University of Auckland
y.koh@auckland.ac.nz

Giovanni Russello

University of Auckland
g.russello@auckland.ac.nz

Danielle Lottridge

University of Auckland
d.lottridge@auckland.ac.nz

Abstract—Phishing is when social engineering is used to deceive a person into sharing sensitive information or downloading malware. Research on phishing susceptibility has focused on personality traits, demographics, and design factors related to the presentation of phishing. There is very little research on how a person’s state of mind might impact outcomes of phishing attacks. We conducted a scenario-based in-lab experiment with 26 participants to examine whether workload affects risky cybersecurity behaviours. Participants were tasked to manage 45 emails for 30 minutes, which included 4 phishing emails. We found that, under high workload, participants had higher physiological arousal and longer fixations, and spent half as much time reading email compared to low workload. There was no main effect for workload on phishing clicking, however a post-hoc analysis revealed that participants were more likely to click on task-relevant phishing emails compared to non-relevant phishing emails during high workload whereas there was no difference during low workload. We discuss the implications of state of mind and attention related to risky cybersecurity behaviour.

I. INTRODUCTION

Phishing is a form of social engineering scam where individuals are convinced to share sensitive information, such as credentials, or to download harmful software. Phishing causes damage to both individuals and organisations, and its prevalence is increasing: Verizon’s 2022 report [27] shows that 36% of data breaches involve phishing, and 82% of breaches involve the human element. Filters and automatic scam detection are used to prevent scams from reaching inboxes, however people can be considered the last line of defense. Their ability to identify a phishing scam directly impacts the success of the attack.

Identifying phishing scams is not straightforward. Typically, people engage with their messages or email inbox in good faith, and focus on content. There may be certain cues to phishing, such as a strange sender address or an unusual URL. If people pay attention to these cues, this

raises suspicion, which then changes people’s mindsets from focusing on content to questioning the email’s legitimacy [32], [4]. Detecting phishing emails requires attention and cognitive effort [21], [8], [28].

Workload impacts task performance. High workload is associated with higher fatigue, and reduces performance in attention-demanding tasks [7]. We hypothesise that under high workload, people’s attention will be focused on the primary task, i.e., dealing with the content of email, leaving little attention for determining legitimacy. Classic models of human attention describe attention as finite, with a limited capacity [15]. Thus, we speculate that when people are more mentally engaged in managing email, this means less attention is available to notice phishing cues, and therefore will result in lowered ability to detect phishing emails.

Research on phishing email susceptibility tends to use one of two methods: realistic phishing campaigns or in-depth surveys [35]. Realistic phishing campaigns have excellent ecological validity because participants are not primed to detect phishing, but with this method it is difficult to know much about the context of the user. Thus, phishing campaign research tends to include variables such as demographics, personality traits, and design aspects of the phishing simulation [35]. Surveys on phishing are able to carefully study attention toward a range of cues, but ecological validity is sacrificed because the participant is primed to look for phishing and it is outside of their usual email task context. Neither of these methods are ideal for studying the effects of attention, workload, and stress on phishing email susceptibility. We are inspired by HCI studies of email management and stress (cf. [1]) to create a scenario-based in-lab experiment. We provide a realistic scenario to participants: to collect information about an event and its budget through a series of emails, within a typical email interface. This method enables us to ensure that all participants experience the same amount of workload, and critically, it enables us to systematically vary workload, and to identify a relationship between workload and phishing detection. This methodological approach prioritises ecological validity as much as possible while controlling environmental factors. It enables us to measure high-quality indicators of experienced workload such as physiological data, and measure relevant cybersecurity interactions such as hovering over

URLs.

We used this approach to answer the research question: *Does workload influence phishing susceptibility?*

This article makes two contributions:

- 1) We used a methodological approach of an email management task within an in-lab experiment to study phishing susceptibility, a method which has not previously been applied to study phishing behaviours. We validate that experienced workload differs between the high and low workload conditions with significant differences by using physiological stress indicators of eye-tracking fixation duration, electrodermal activity (EDA) and self-reported stress. We demonstrate that email management behaviour changes with workload, with measures such as email reading time differing significantly between high and low workload.
- 2) We compare the risky cybersecurity behaviours of clicking on links and attachments between the high and low workload conditions, and we observe some expected and some unexpected results: under high workload people did spend less time on emails, but they did not hover over links more. Most importantly, under high workload, participants did not click on phishing links more. In post hoc analysis, we identified possible reasons for this unexpected behaviour that may lead to understanding how to reduce the effectiveness of phishing attacks.

II. RELATED WORK

Training is a popular way of educating people about phishing prevention and detection [2], [19]. Periodic retraining is thought to ensure awareness of phishing does not fade [13], [22]. Yet, phishing-related knowledge has a limited effect on phishing detection because it requires elaboration [29]. Vishwanath et al. point out that users fall for phishing because they do not pay enough attention to processing the email [29]. A phishing campaign study run in a financial institution found that those who took more elaborate care in evaluating email legitimacy were less likely to click on a phishing email [3].

Attention is a limited resource, and when reading emails, people tend to focus their attention on understanding the email message content, and devoting attention to verifying the email's legitimacy is typically secondary. When reading an email, people only shift their mindset to investigate its legitimacy if they notice something abnormal or suspicious [32]. This process would be explained by the Kahneman and Tversky's concept of the human mind as a dual process system [16]. This theory proposes that the human mind has two different processing systems, one is fast and effortless, and is based on intuitions and heuristics (System 1); the other is slow and effortful, and it requires systematic reasoning (System 2). This theory has gained popularity in phishing research in recent decades [28], [30], [9], [33]. Processing email content requires attention, and consideration of legitimacy would normally involve System 1 thinking and so in the absence of anything unusual, legitimacy is not questioned. Only when something does appear unusual would System 2 become engaged.

Our research is on the effects of workload on phishing susceptibility, because high workload leads to reduced work

performance [20], [18]. Studies in both car driving [18] and flight simulation [20] contexts have shown that as the environment becomes more complex, workload increases, leading to more errors and decreased decision-making efficiency. Similarly, when people are asked to process a large amount of emails in a limited amount of time, the users' workload would increase. Previous work has shown that such workload and time pressure can influence the care with which individuals process emails [23].

Previous research on the impact of email workload on phishing susceptibility has had a variety of results. People with higher email loads are more likely to habitually respond to emails, and are thereby more likely to be deceived by phishing emails with urgency cues [29]. A simulated phishing campaign in a healthcare setting found that workload has a significant relationship with the likelihood of clicking on phishing links [12]. An email management study found that higher email load did not affect the ability to classify or respond to normal emails, even though the high email load condition was deemed as more challenging [24]. However, one simulated phishing study found that people with high cognitive loads were less susceptible to phishing — when that cognitive load was about detecting legitimacy [21].

These findings further motivate our investigation into the impact of workload on susceptibility to phishing. We hypothesise that under high workload, individuals allocate more attention to evaluating and responding to normal emails, leaving even less attention to verify their legitimacy compared to situations with lower workloads. We therefore anticipate that individuals under high workload are less likely to detect phishing emails and so are more susceptible to phishing attacks.

III. RESEARCH GOAL AND HYPOTHESES

Based on the literature, we suspect that workload will influence phishing susceptibility.

To investigate how interaction with phishing emails differs with varying workloads, we focus on three behavioural outcomes: reading time, hovering over phishing URLs, and clicking phishing links/attachments, during low and high workload conditions. We expect that people will spend less time on each email during the high workload compared to the low workload because there will be more tasks to do and more emails to look at under high workload. Similarly, due to the high workload, we expect that people are less likely to hover over links when under high workload because this action is mainly for security checks instead of helping them complete their primary tasks. Finally, as a result of less time spent reading carefully and fewer checking behaviours under high workload, we expect that participants will click on more phishing links/attachments, compared to when under low workload.

H1: Participants will spend less time reading each email under high workload compared with under low workload.

H2: Participants will hover over links less when under high workload compared with under low workload.

H3: Participants will click on phishing links/attachments more under high workload compared with under low workload.

IV. METHODOLOGY

To attempt what ecological validity was possible while also retaining precise control over workload, we devised an email management application that presents a scenario, tasks, and information through a series of emails delivered throughout the session, while also capturing behavioural data. The scenario refers to the participant as an employee who is tasked with a typical event management task. Providing a scenario rather than a sole primary task enables us to introduce an implicit secondary task, which is to manage email unrelated to the primary task. Thus, the primary task is to read event emails and fill in spreadsheets with information contained in the emails, and the secondary task is to consider any other email in the inbox. With a scenario-based approach, we were able to expose participants to phishing emails without explicitly priming them, thereby avoiding bias in studying their phishing susceptibility. Further, we use this controlled in-lab context to vary workload in the exact same way for each participant, and to collect precise behavioural measures. Taken together, this approach allows us to answer our research question: *Does workload influence phishing susceptibility?*

A. Scenario and Instructions

Participants were instructed to imagine that they were a temporary office worker, taking the place of a sick staff member organising an event (see Appendix A for full scenario and instructions). They were asked to look at the staff member’s email inbox, and process all emails as if the emails were sent to them. This involved filling in an event planning spreadsheet based on information in the emails. To simulate a realistic email inbox, the email inbox in the study was designed to contain not only event-related emails but also other internal and external emails. We explained to the participants that there were different types of emails in the inbox and they should highlight emails that they think are important, and mentioned they should report any suspicious emails. We demonstrated the email interface, and briefly showed them the types of emails they would receive, and the tasks they needed to complete. We feel this introduction resembled the kind of brief training that would be provided to a temporary office worker hired for a short-term specific task.

B. Workload Conditions

The experiment used a within-subjects design, meaning that all participants experienced both the low workload condition and the high workload condition. The ordering of the workload conditions was counterbalanced across participants. Each of the high and low workload sections was exactly 15 minutes in length. The two workload sections differed in the number of emails (30 emails in high, 15 in low; Table I) and consequently the amount of work that needed to be done. Under high workload, the participants were required to complete twice the amount of work as compared to the low workload, in the same amount of time. We designed and pilot tested the workload conditions so that the participants were able to complete the low workload section on time, and would find it difficult in the high workload section.

TABLE I. THE NUMBER OF EMAILS USED IN THE LOW AND HIGH WORKLOAD SESSION

Emails	Low workload	High workload
Phishing emails	2	2
Event-related emails	5	10
Non-relevant emails (ads, internal, external emails)	8	18
Total	15	30

C. Phishing Email Design

The four simulated phishing emails mixed relevance and type of phishing attack, i.e., link and attachments. Thus, we include one task-relevant link attack, one less-relevant link attack, one task-relevant attachment attack, and one less-relevant attachment attack (see Appendix A). When we say “task-relevant”, we do *not* mean “spear-phishing”, where the emails utilise specific knowledge the attackers have about the user or their work. Rather, our “task relevant” emails simply used generic terms (e.g., report, document) that might plausibly relate to the user task. All four phishing emails were actionable, leading to potential danger. The sender email address and link URLs in each phishing email had unusual or mangled versions of well-known domains, and with no obvious reasonable connection to the user task.¹ For example, the scenario was that the event was for people coming from institutions in our region, and the phishing email sender addresses, URLs, and content had no connection to our region.

D. Email Management Application

We developed a custom application for displaying emails, allowing the participants to complete the primary task of the study, and measuring behavioural aspects such as reading time, hovering and clicks, as well as replies and reporting. Our email client design was based on the Gmail interface (Figure 1), with a list of emails displayed on the left and the email content displayed on the right. Below the email client are tables that participants fill in for the work-task. This bottom section simulates an event planning spreadsheet. Participants are asked to extract information from emails to complete the spreadsheet. Both buttons in the interface are interactive: a report button and a highlight button. While we are only interested in the reporting, we include another button because having only the report button might prime participants to focus more on security, and therefore bias their behaviour. The icons we used for the buttons are the same as Gmail, so the participants’ knowledge about these button functionalities can be transferred from using Gmail to our application.

As described in the next section, we use an eye tracker to measure which part of the interface participants are attending to. To support accurate eye tracking with stable “areas of interest (AOI)”, we designed our application to stay fixed within the left side of the screen. A browser can be opened on the right side of the screen for viewing embedded links and PDF attachments. Both kinds of phishing attacks could be opened in a browser, i.e., URLs and PDFs. We selected the PDF format to minimise the need to switch between different

¹Senders: hayley235sd@outlook-office.co, nginx@vmi398623.contaboserver.net, jay.chapman@tfac.or.th, and quesisteam@kuqpw23.or.th.online. URLs: <https://139hf.trk.elasticemail.com/tracking/click?...> and <http://d0cs.g00gle.online/...>

applications. The use of the eye tracker meant we needed to place all elements on screen without overlapping, therefore we were not able to simulate the default browser Gmail layout, i.e., different browser tabs or pages. For simplicity, we did not include the Gmail side menu bar for selecting different inboxes.

E. Physiological Metrics

All participants carried out the study in the same environment, with the same equipment and workload conditions. We use several physiological sensors to assess cognitive load. Eye movement has been found to be an indicator of cognitive load [5], [31], [26], where high cognitive load is associated with lower fixation rates and higher fixation durations. Electrodermal activity (EDA) is another good indicator of cognitive load in terms of physiological arousal and stress [11], [25], [17], with more EDA peaks associated to greater emotional intensity, stress, and higher cognitive load [25], [6]. The Tobii Pro X3-120 eye tracker was used for recording eye movements and gaze location on screen at 120 Hz. In our pre-processing, we discarded fixation durations of less than 60 ms and used a moving window average of 3 samples to reduce the noise in the data. The Empatica E4 wristband was used for recording the participant's EDA. We calculated the number of emotional arousal peaks triggered in each session to estimate the participants' cognitive load. Consistent electrodermal data was difficult to obtain from some participants, so data from only 14 participants was included in the analysis.

F. Post-study Interview and Questionnaire

To better understand participants' interactions with the phishing emails in their inbox and their attitude toward the study, we conducted a short interview after the experiment. This was followed by administering the NASA TLX questionnaire to obtain subjective estimate of the workload through six measurements: mental demand, physical demand, temporal demand, performance, effort, and frustration [10]. We used a 10-step scale for each of the measurements, and used the sum to approximate participant's overall subjective workload. Participants completed the NASA TLX questionnaire twice, to reflect their experience in the first and second workload sessions (both TLX questionnaires were completed after the experiment).

In the post-study interview, we presented the four phishing emails to the participant one by one and asked the following questions: "What do you think about this email? Is there anything you want to tell me about when you read this email?" The open-ended questions were used to collect the participants' perspectives on the phishing emails, because we wanted to assess their reactions while biasing them as little as possible. We then showed the participants the list of emails they reported and asked for reasons for reporting. At the end, we collected participants' feedback on the study and their personal email reading habits.

G. Participants

We recruited participants with basic technology knowledge. In all, 26 participants were recruited, including 13 participants from an undergraduate course, 10 graduate students from

technology, engineering and arts, and 3 administrative staff. Among these participants, 16 were female and 10 were male. 16 of them were from 18 to 24 years old, 8 were from 25 to 34 years old, and 2 were 45 or older.

We acknowledge limitations on generality in this recruitment, but suggest that it might well match our study scenario, people hired for temporary office work using email to help organise a conference. We argue that these participants are suited for the scenario as some of the emails used in the scenario are similar to what they will receive in their real-life, and the primary task was designed so that they have no difficulty completing it. Also, our study had a within-subject design which would help address some issues with the sample and sample size.

Participants were given \$20 shopping vouchers for their participation. Our study protocol was reviewed and approved by the Human Participants Ethics Committee of the university. At the end of each session, we debriefed the participant on the true nature of the study (i.e., that we were studying phishing susceptibility).

V. RESULTS

This section begins with an assessment of the effectiveness of our workload conditions, then proceeds to test hypotheses, and finishes with post hoc analyses to further explore the results.

In order to confirm that our manipulation of workload was effective, we assess participants' cognitive load through objective and subjective measures. The objective measures included the number of EDA peaks from the wristband, and the median fixation duration and fixation rate from the eye tracker. We then performed one-sided, paired t-tests on each of these features. EDA peaks and median fixation duration show significant differences between low and high workloads (see Table II). There were significantly more EDA peaks and higher median fixation duration during the high workload compared with the low workload, indicating that the participants were experiencing more cognitive load. It is worth noting that even though the fixation rate did not result in significant differences, the direction of the result matched the literature (lower fixation rate suggests a higher cognitive load). In addition to these objective measures, we also observed that 14 out of the 26 participants finished the low workload session at least one minute early, whereas only three participants finished the high workload session early. This also suggests that participants were able to manage the workload in the low workload session, but struggled in the high workload session. To assess subjective workload ratings between the low and high workload conditions, a one-sided, paired t-test was performed. Participants reported that they were experiencing significantly more mental workload during the high workload session compared with the low workload session $t(25) = -4.87, p < 0.001$. These physiological, behavioural, and self-report results show that the sessions effectively caused low and high workload as intended.

A. H1: Email Reading Time

We hypothesised that participants would spend less time reading each email under high workload. We therefore examined the time spent reading each email, and also the percentage of emails read.

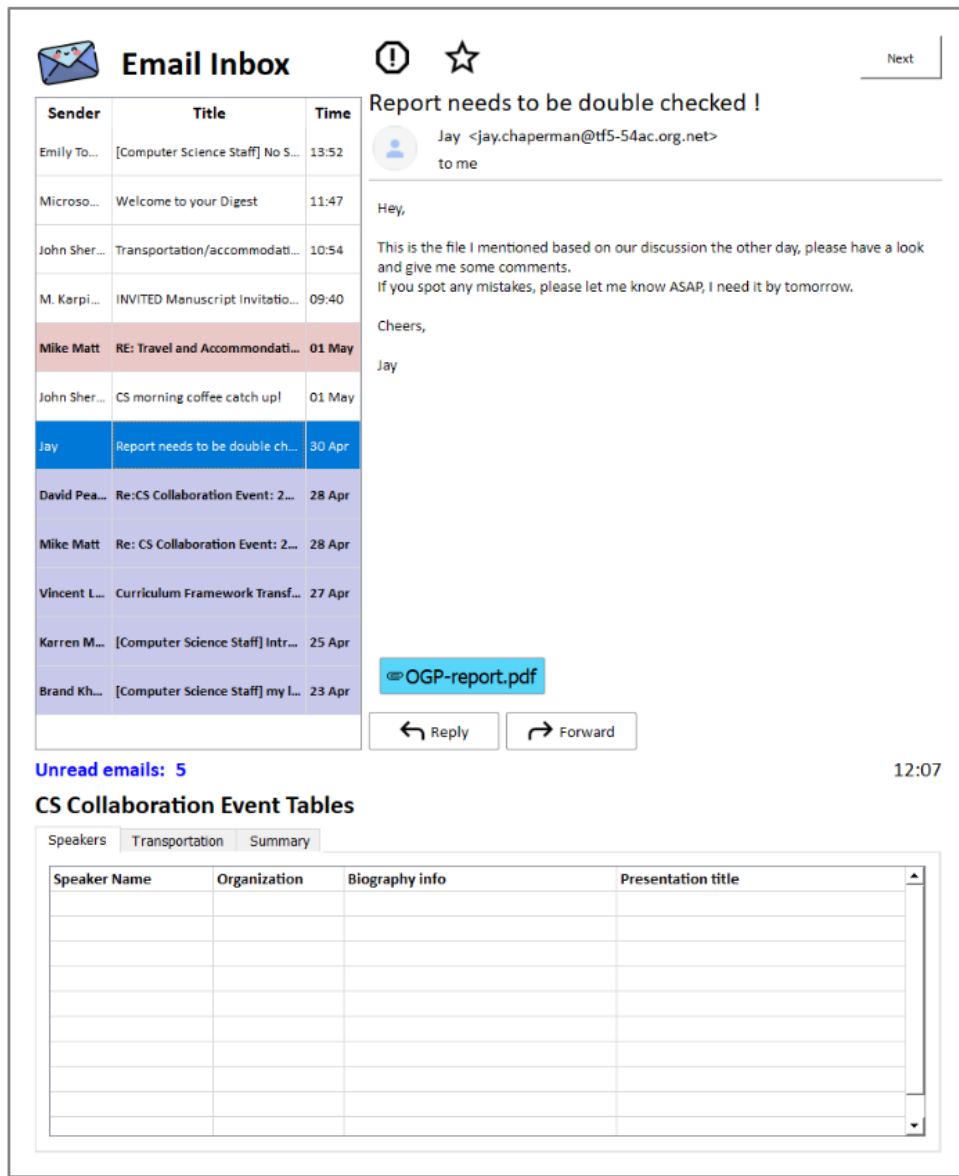


Fig. 1. The user interface of the email management application. Under the ‘Email Inbox’ heading in the top left area, the list of emails is shown, with read emails in white, highlighted in pink, selected in blue and unread in purple. The area in the top right displays the reporting button and the star/highlight button, with the contents of the selected email below. The bottom part of the screen is the task spreadsheet, where participants input information.

TABLE II. OBJECTIVE AND SUBJECTIVE MEASURES OF COGNITIVE LOAD BETWEEN LOW AND HIGH WORKLOADS. EFFECT SIZES MARKED N.S. INDICATE TEST WAS NOT SIGNIFICANT.

	Low workload		High workload		t	p	Cohen's d
	Mean	SD	Mean	SD			
EDA peaks (per min)	3.63	2.18	5.26	2.65	-3.55	0.002	-0.67
Med fixation duration (ms)	166.17	27.33	191.36	23.08	2.50	0.01	-0.19
Fixation rate (per sec)	2.99	0.46	2.94	0.42	1.00	0.328	n.s. 0.12
NASA TLX	30.58	6.66	39.19	7.21	-4.87	<0.001	-1.24

To test the hypothesis, we performed a t-test on the participants’ average email reading time. The test was paired,

because of our within-subjects design, and one-sided to reflect our hypothesis that the high workload condition would show less time. For this, and all other tests we conduct, we used an alpha value of 0.05.

We found that participants did indeed spent less time reading each email under high workload ($M=11.44$, $SD=4.75$) than under low workload ($M=21.26$, $SD=9.69$), and the result was significant $t(25) = 6.36$, $p < 0.001$, with a large effect size ($Cohen's d = 1.286$).

We then compared the percentage of unread emails in the high and low workload conditions, and performed a paired, one-sided t-test. We found that participants read a significantly smaller proportion of emails in the high workload compared with the low workload ($t(25) = -2.14$, $p = 0.02$).

This evidence supports H1.

B. H2: Link Hovering Behaviour

We hypothesised that participants would hover over links in emails less in the high workload condition. This is important because the links were hidden by a button or hypertext, so hovering over links to reveal the URL is the way to assess whether the link is suspicious in order to decide whether or not to click.

We analyzed the proportion of emails where participants hovered over links, as well as the average time hovering over the links. There are embedded links in all categories of the emails.

Considering all categories of email, the proportion of emails where the user hovered was similar in the high workload condition ($M = 0.14, SD = 0.06$) to that in the low workload condition ($M = 0.09, SD = 0.09$), with no significant difference found with a paired, one-sided t-test ($t(25) = 2.020, p = 0.973$). Also, across all categories of email, the amount of time when hovering was also similar between low workload ($M = 13.67, SD = 28.16$) and high workload ($M = 12.58, SD = 35.82$), again with no significance with a one-sided t-test ($t(18) = 0.216, p = 0.584$). (Note that pairwise tests exclude a few participants who never hovered — but a non-paired test gave a similar result.)

Therefore, H2 is not supported. We have no evidence that people hover over fewer emails under high workload, or hover for less time.

C. H3: Phishing Email Clicks

We hypothesised that more people would click on phishing links in the high workload condition.

To test this hypothesis, and considering our within-subjects design, we conducted a McNemar test which compares the behaviour of the same people in each condition. Table III shows the results. There was no significant association between workload conditions and phishing clicks, $\chi^2(1, N = 26) = 0.75, p = 0.386$. Moreover, we note that more people clicked in the low workload condition and did not click in the high workload condition, rather than the reverse.

The McNemar test considers whether each participant clicks on a phishing link in each of the two conditions, whereas some participants might click a link in one of the two phishing emails in the condition, and others might click on both. To address this issue, we also conducted a one-sided Wilcoxon Signed Rank test (similar to a paired t-test but non-parametric because click numbers will simply be zero, one, or two). The result was again not significant $V = 79.00, p = 0.078$.

We conclude that H3 is not supported.

D. Post Hoc Analyses

We conducted post hoc analyses to look into several aspects of emails to further understand our results, in particular why it seems that there was more phishing clicks during low workload compared to high workload. This unexpected result led to more extensive post hoc analysis than we would normally conduct.

TABLE III. CO-OCCURRENCE MATRIX OF PARTICIPANTS’ PHISHING EMAIL CLICKING BEHAVIOUR UNDER LOW AND HIGH WORKLOADS. TOP LEFT IS THE NUMBER WHO CLICKED ON AT LEAST ONE LINK IN BOTH LOW AND HIGH WORKLOAD CONDITIONS, BOTTOM RIGHT IS NEITHER, OTHER CELLS SHOW NUMBER WHO CLICKED IN ONE CONDITION BUT NOT THE OTHER.

H3	Low Clicked	Low Not Clicked
High Clicked	4	4
High Not Clicked	8	10

The tests reported in this section are exploratory only, and we do not make corrections for multiple tests. We therefore need to be cautious in interpreting the results, but they may suggest where future studies might focus.

1) *Ordering effect*: As a participant progresses through the conditions in our study, we can expect that they learn more as they go, and may come to better understand the task and the system by the end of the experiment. This therefore might create an effect of the order in which they experienced the two conditions. We investigate whether the order of workload sessions had an effect on the participants’ behaviour or on their workload. We explored whether the order of low and high workload sessions might affect user behaviour significantly, but found no evidence for EDA, fixation duration, NASA TLX, email reading time, hovering, or click rate.

2) *The impact of workload and the task-relevance of phishing emails on clicking behaviour*: Our experiment included two task-relevant phishing emails and two less-relevant phishing emails. The participants would see one potentially task-relevant phishing email and one less-relevant phishing email during each workload condition. To investigate the effect of phishing email task-relevance and clicking, we conducted a McNemar test (Table IV), comparing the number of participants who clicked on emails with the number who did not (including those who did not open the email). The result was significant: participants were more likely to click on phishing emails when the email was more relevant to the task context $\chi^2(1, N = 26) = 6.75, p = 0.009$.

TABLE IV. NUMBER OF CLICKS ON TASK-RELEVANT AND LESS-RELEVANT PHISHING EMAILS

Rel. All Workloads	More Rel. Clicked	More Rel. Not Clicked
Less Rel. Clicked	4	1
Less Rel. Not Clicked	11	10

To investigate whether this relationship applied to both the low and high workload sessions, we analyzed participant’s clicking behaviour in low and high workload separately (Table V). The McNemar test result showed that the relationship was significant in the high workload, $\chi^2(1, N = 26) = 6.12, p = 0.013$, but insignificant in the low workload, $\chi^2(1, N = 26) = 3.12, p = 0.077$. In other words, participants were significantly more likely to click on task-relevant phishing emails compared to non-relevant phishing emails under high workload, whereas there was no difference in the low-workload condition.

3) *The impact of task-relevance on email reading time*: Our research found that participants spent significantly less time reading emails during high workload, but this did not result in more phishing email clicks. To further investigate how

TABLE V. NUMBER OF CLICKS ON TASK-RELEVANT AND LESS-RELEVANT PHISHING EMAILS DURING LOW AND HIGH WORKLOAD CONDITIONS

Rel. Low Workloads	More Rel. Clicked	More Rel. Not Clicked
Less Rel. Clicked	4	1
Less Rel. Not Clicked	7	14

Rel. High Workloads	More Rel. Clicked	More Rel. Not Clicked
Less Rel. Clicked	0	0
Less Rel. Not Clicked	8	18

participants interacted with email, we investigated reading time of emails across three categories: relevant legitimate emails, non-relevant legitimate emails, and phishing emails; and those categories in low and high workload conditions. These times are shown in Table VI. We conducted a two-way repeated-measures ANOVA on this data, testing how the reading was affected by the category of email, and the workload condition. The test showed an interaction effect between category and condition: $F(2, 144) = 6.30$, $p = .002$, $\eta_p^2 = .08$. There were main effects for both category ($p < .001$, $\eta_p^2 = .34$) and condition ($p < .001$, $\eta_p^2 = .09$). Exploration of the differences showed all low workload categories to show longer times than high workload categories, and relevance shows the largest differences over the other categories. Figure 2 illustrates the differences. In this analysis we did not distinguish between more and less relevant phishing emails because there were only two of each kind. However, our analysis in the section above showed that it was the more relevant emails that get more attention in the high workload condition.

TABLE VI. PARTICIPANTS' AVERAGE EMAIL READING TIME (SECS) OF DIFFERENT CATEGORIES OF EMAILS BETWEEN LOW AND HIGH WORKLOADS

	High		Low	
	M	SD	M	SD
nonrel	4.81	2.42	10.79	7.01
phish	4.99	4.52	6.03	3.48
rel	21.59	10.03	43.13	23.00

4) *Attention to Subject, Sender and Body*: We were interested to explore whether workload impacted attention to phishing cues. We use the eye-tracking data to look at three areas of interest (AOIs): subject line, email sender information, and email body. Our primary interest is the sender information, because it could be used to determine legitimacy.

Table VII shows descriptive data for gaze time on the sender AOI across different categories of emails, and Figure 3 illustrates the differences. We again conducted a two-way repeated-measures ANOVA on this data, testing how the reading was affected by the category of email, and the workload condition. The test showed no interaction effect between category and condition: $F(2, 144) = 2.94$, $p = .056$, $\eta_p^2 = .04$. There was also no main effect for both category ($p = .151$, $\eta_p^2 = .03$) or condition ($p = .053$, $\eta_p^2 = .03$).

It seems there was little difference in how long participants looked at the sender information. In particular, the sender information in the phishing emails did not receive the lengthy gaze that might be expected. Examining just the phishing emails, we found no significant difference between low and high workload conditions. We also found no evidence of a

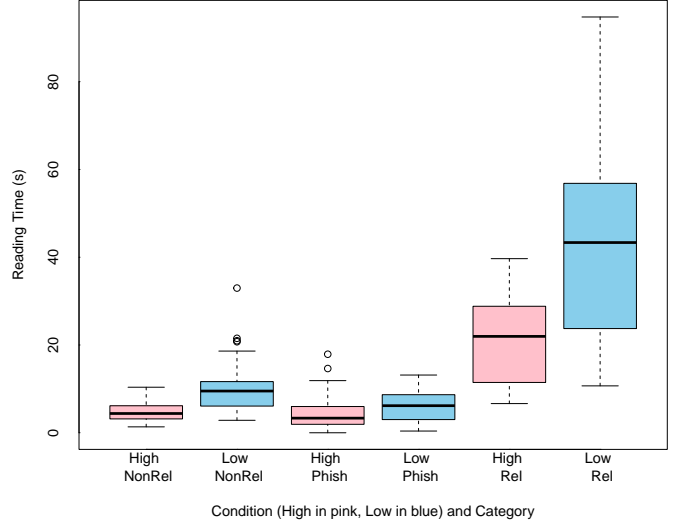


Fig. 2. Boxplot of email reading time by workload condition and email category. Coloured box is central quartiles, whiskers are outer quartiles, and circles are outliers.

relationship between gaze time on the sender and clicking behaviour.

The email subject gaze data also showed no significant difference between categories or conditions. The gaze data for the email body resembled the reading time, discussed earlier.

TABLE VII. PARTICIPANTS' SENDER GAZE READING TIME (SECS) OF DIFFERENT CATEGORIES OF EMAILS BETWEEN LOW AND HIGH WORKLOADS

	High		Low	
	M	SD	M	SD
nonrel	0.71	0.51	1.64	1.62
phish	1.30	1.55	1.07	1.03
rel	1.70	1.18	2.64	1.55

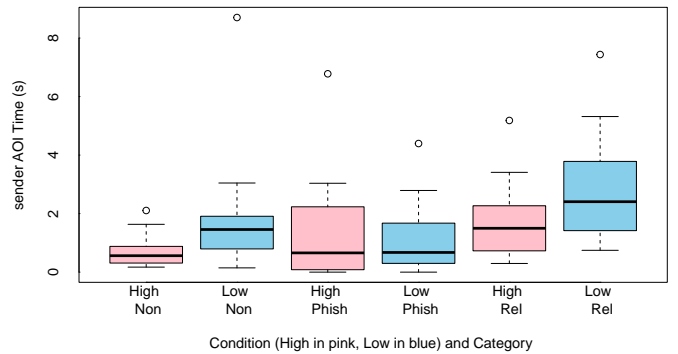


Fig. 3. Boxplot of sender gaze time by workload condition and email category. Coloured box is central quartiles, whiskers are outer quartiles, and circles are outliers.

5) *Link vs. attachment phishing emails*: The four phishing emails used in our study can also be categorised into two

phishing emails that contain phishing links and two emails that contain phishing attachments. In each workload session, the participants would receive one phishing email with a phishing link and one phishing email with an attachment. To investigate whether there was a difference in clicking behaviour when seeing different types of phishing emails, we performed McNemar tests and found no significant differences for low workload ($\chi^2(1, N = 26) = 0.75, p = 0.386$) or high workload ($\chi^2(1, N = 26) = 0.12, p = 0.724$).

6) *Reporting behaviours*: In our study, we provided a button for the participants to report any emails that they feel suspicious. 18 out of the 26 participants reported at least one email in our study, resulting in 24 unique emails being reported. We decided to focus our analysis on the 15 emails that were reported at least twice. Among these emails, all of them are either non-relevant legitimate emails or phishing emails. The top five most reported emails were commercial advertisement emails (top four, reported at least six times) and phishing emails (less-relevant phishing emails, reported five times). When we asked the participants for their reasons of reporting, the main reason for reporting was an imprecise feeling that the email was suspicious. The prevalence of advertisement emails being reported suggests to us that people may have a conception of “suspicious” that differs from phishing.

7) *Interviews*: After the email processing session, we had a short interview session asking users’ perception about the four phishing emails. The data was analysed through open coding to group and label user responses. Half of the participants mentioned that when processing emails, they would first skim through the email to extract important content, in order to decide whether the email is relevant to their task, then decide whether they would spend effort responding to the email. Regarding the participants interaction with the phishing emails, four participants mentioned they would click on any links or attachments in the email without much thinking. Their rationale was to use the information in the landing page and attachments to help them understand the email message and determine legitimacy: a risky practice.

Overall, 14 out of the 26 participants mentioned that at least three of the four phishing emails were not relevant, and so they ignored the emails. This observation may explain why we observed a low phishing link/attachment click rate in the study. The most common reason for participants claiming the phishing emails looked suspicious was sender address, noted by seven participants. Following this, suspicious subject, vague content, and visual presentation of the email were also mentioned by two participants each. In general, participants appeared uncertain how to assess legitimacy with rigour.

VI. DISCUSSION

A. Method

Our methodological approach to the study of phishing susceptibility was a scenario-based, in-lab experiment, where we successfully manipulated workload. Previous research on phishing has used a simulated email client [14] as context for participants to make explicit judgements of whether each email was phishing or legitimate, rather than to facilitate a primary email management task with an implicit secondary task of

considering non-task relevant email. Being in-lab meant that we could measure eye-tracking and EDA for precise measures of participants’ mental workload. We were able to collect precise analytics related to reading time and hovering. Phishing campaigns are the most popular study method because they maximise ecological validity, where people are not just reading emails, but also doing complex and overlapping sets of realistic tasks. Our approach simulates this real-world scenario by giving participants several tasks to do. Our approach provides another option in the balance of ecological validity and control over the environment and precise measurement. We were able to use this approach to offer insights into how email reading differs significantly across low and high workload, and implications this has for phishing susceptibility. It is worth noting that the ecological validity of the study strongly depends on how relevant the scenario is to the participants’ real-world experience. It is possible to create a scenario that closely simulates the types of jobs participants have in the real world. However, customising the scenario for each participant would increase the cost of conducting the study and make comparisons between participants less feasible.

B. Workload and Relevance

Before conducting the study, we believed that when under high workload, users would spend less time reading each email and thus be less careful with potentially risky clicking behaviours. The first part of our premise was validated (i.e. lower average reading during high workload) however this did not result in significantly fewer hovers over links, and more dangerous clicks. Instead, we found there were more clicks under low workload (though the analysis was not statistically significant). The most interesting significant relationship we found in post hoc exploration was between the potential relevance of phishing emails and workload, where relevance matters when workload is high; people are more likely to click on relevant phishing emails under high workload, whereas there is no association when under low workload.

During the low workload, we observed that when participants had less work to do, they tend to spend more time on even non-relevant emails (including phishing emails). When people have more time to read less-relevant emails, they may be at risk of paying more attention to potential phishing emails. This potential relationship should be followed up in future work. On the other hand, during high workload, users tended to interact with emails differently. Compared with low workload, they processed relevant emails twice as quickly as when in low workload, and they tended to ignore or rush through non-relevant emails. Under high workload, even when phishing emails are well-crafted, if the content does not seem relevant, people seem likely to ignore it. The interview results also confirm that, many participants classified the phishing emails as non-relevant emails, which resulted in ignoring the emails. However, ignoring phishing emails in such conditions does not necessarily mean people are safe from the attack. People may decide to return to the emails later. In our study, two participants reported that if they had time, they would have returned to some emails to look at them, which might have potentially led to clicking on the phishing emails. In fact, we did observe one participant who finished the primary task a few minutes early (in the low workload), then went back to

reading less relevant emails, and clicked on links in phishing emails.

Our research established that participants had more physiological arousal during high workload, and were under pressure to read more quickly, which may influence how they react to seemingly relevant email content. This potential relationship between relevance, physiological stress, workload and phishing susceptibility should be further investigated in future work.

Our post hoc result that people are more likely to click on potentially relevant phishing emails under high workload is related to the phenomena of spear phishing. However, spear phishing normally implies knowledge specific to the recipient. The more relevant emails in our study had fairly generic terms (“report”, “document”) that were successful, meaning that people in large organisations might be at risk of phishing emails with generic terms relevant to the organisation.

C. Need for Support

The phenomenon seen in the low workload condition is also a concern. Where people had more time, they often engaged more with phishing emails, even those with no relevant connection to their task. In some cases, they clicked on dangerous links. This suggests that they need more support in determining the legitimacy of emails.

In our post-study interviews, we showed the participants our phishing emails and asked for comments. Even though some participants mentioned they paid attention to the phishing email sender, none of them mentioned checking the actual URL. Most participants commented on the email content and visual presentation. And although the email sender address is very helpful in determining the legitimacy of the email, embedded phishing links and attachments are the entry point of the attack.

We believe that research should focus more on designing interventions to make it easier for users to determine email legitimacy. For example, they might be helped by making more visible any URLs hidden by buttons or text. Moreover, it should be easier for users to determine the provenance or domain names used in links and sender email addresses. Users spend much time examining the body of phishing emails, where the content and presentation are under the control of attackers, and can seem professional, convincing, and seem urgent and offer malicious “calls to action”. Users need support in noticing such calls for action, but then they still need support in accessing more reliable information to determine email legitimacy.

There have been designs and studies of security tools to improve phishing detection [34], and there is a need for more research of that kind. Not only should security tools or email plug-ins contribute to improving phishing detection, but such tools or email plug-ins should alert users to cues, and then facilitate careful examination of reliable details. Recalling Kahneman’s model, there should be support for users to transition from heuristic System 1, to rigorous System 2, and then help with information and guidance for System 2.

D. Limitations

In our study, we base our findings on users’ clicking behaviours on phishing links or attachments. We acknowledge

that clicking on phishing links does not necessarily lead to successful phishing. For credential harvesting phishing attacks, users can choose not to enter their credentials once they visit the landing page. However, as long as the users open the phishing landing page, tracking in the links can allow attackers to identify users for further malicious attempts.

A few participants reported that their interactions with our system did not reflect their typical email reading behaviours. We summarise their explanations: 1) the types of emails used in the study were different from their own email inbox, 2) the time pressure led to a more task-focused approach, 3) participants could report any suspicious emails, but most do not have the habit of reporting emails.

The generalisability of our findings are limited due to our small sample, toward participants having some knowledge related to technology. However, our within subjects study design allows like-for-like comparisons, the sample did reflect people who might do the kind of temporary office work in our scenario.

VII. CONCLUSIONS

In this research, we explored the relationship between workload and users’ phishing susceptibility. Our hypothesis that users are more likely to click on phishing emails under high workload was not supported. Under low workload, participants spent more time to look at each email, which may have led to increased interactions with potentially malicious emails. Conversely, when participants were under high workload, they focused on more task-relevant email and often ignored non-relevant emails, including generic phishing attempts. Post hoc exploration suggested that relevance is a key reason that people engage more with emails, and with low workload they are more likely to click on malicious links. This deserves more study.

Our study was conducted in a lab setting, but simulating a realistic scenario of working on typical administrative tasks and encountering phishing emails along the way. This methodological approach allowed environmental control, precise workload manipulation, allowed collecting valuable physiological data, while retaining some ecological validity. Further investigations could explore the relationships between other situational characteristics and phishing susceptibility, enabling actionable insights of how workplaces can enhance support for phishing detection.

Our initial speculation had been that high workload was related to phishing susceptibility. We now speculate that more important issues are phishing email relevance, and a lack of support for users to reliably determine legitimacy.

ACKNOWLEDGMENTS

We thank the research participants for taking the time to participate in this study. We also thank the reviewers for their valuable inputs and suggestions. Robert Biddle acknowledges funding from the Natural Sciences and Engineering Research Council of Canada (NSERC): RGPIN-2022-04887.

REFERENCES

- [1] F. Akbar, A. E. Bayraktaroglu, P. Buddhharaju, D. R. Da Cunha Silva, G. Gao, T. Grover, R. Gutierrez-Osuna, N. C. Jones, G. Mark, I. Pavlidis *et al.*, “Email makes you sweat: Examining email interruptions and stress using thermal imaging,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–14.
- [2] A. Baillon, J. De Bruin, A. Emirmahmutoglu, E. Van De Veer, and B. Van Dijk, “Informing, simulating experience, or both: A field experiment on phishing risks,” *PLoS one*, vol. 14, no. 12, p. e0224216, 2019.
- [3] J. Buckley, D. Lottridge, J. Murphy, and P. Corballis, “Indicators of employee phishing email behaviours: Intuition, elaboration, attention, and email typology,” *International Journal of Human-Computer Studies*, vol. 172, p. 102996, 2023.
- [4] C. I. Canfield, B. Fischhoff, and A. Davis, “Quantifying phishing susceptibility for detection and behavior decisions,” *Human factors*, vol. 58, no. 8, pp. 1158–1172, 2016.
- [5] S. Chen, J. Epps, N. Ruiz, and F. Chen, “Eye activity as a measure of human mental effort in hci,” in *Proceedings of the 16th international conference on Intelligent user interfaces*, 2011, pp. 315–318.
- [6] D. Conway, I. Dick, Z. Li, Y. Wang, and F. Chen, “The effect of stress on cognitive load measurement,” in *Human-Computer Interaction—INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part IV 14*. Springer, 2013, pp. 659–666.
- [7] J. Fan and A. P. Smith, “The impact of workload and fatigue on performance,” in *Human Mental Workload: Models and Applications: First International Symposium, H-WORKLOAD 2017, Dublin, Ireland, June 28-30, 2017, Revised Selected Papers 1*. Springer, 2017, pp. 90–105.
- [8] B. Harrison, E. Svetieva, and A. Vishwanath, “Individual processing of phishing emails,” *Online Information Review*, 2016.
- [9] B. Harrison, A. Vishwanath, Y. J. Ng, and R. Rao, “Examining the impact of presence on individual phishing victimization,” in *2015 48th Hawaii International Conference on System Sciences*. IEEE, 2015, pp. 3483–3489.
- [10] S. G. Hart and L. E. Staveland, “Development of nasa-tlx (task load index): Results of empirical and theoretical research,” in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.
- [11] J. A. Healey and R. W. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.
- [12] M. S. Jalali, M. Bruckes, D. Westmattmann, and G. Schewe, “Why employees (still) click on phishing links: investigation in hospitals,” *Journal of Medical Internet Research*, vol. 22, no. 1, p. e16775, 2020.
- [13] D. Jampen, G. Gür, T. Sutter, and B. Tellenbach, “Don’t click: towards an effective anti-phishing training. a comparative literature review,” *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–41, 2020.
- [14] A. Jayatilaka, N. A. G. Arachchilage, and M. A. Babar, “Falling for phishing: An empirical investigation into people’s email response behaviors,” *arXiv preprint arXiv:2108.04766*, 2021.
- [15] D. Kahneman, *Attention and effort*. Citeseer, 1973, vol. 1063.
- [16] —, *Thinking, fast and slow*. macmillan, 2011.
- [17] K. Kalimeri and C. Saitis, “Exploring multimodal biosignal features for stress detection during indoor mobility,” in *Proceedings of the 18th ACM international conference on multimodal interaction*, 2016, pp. 53–60.
- [18] B. H. Kantowitz, “Attention and mental workload,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 21. SAGE Publications Sage CA: Los Angeles, CA, 2000, pp. 3–456.
- [19] D. Lain, K. Kostiaainen, and S. Capkun, “Phishing in organizations: Findings from a large-scale and long-term study,” *arXiv preprint arXiv:2112.07498*, 2021.
- [20] W.-C. Li, F.-C. Chiu, Y.-s. Kuo, and K.-J. Wu, “The investigation of visual attention and workload by experts and novices in the cockpit,” in *Engineering Psychology and Cognitive Ergonomics. Applications and Services: 10th International Conference, EPCE 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part II 10*. Springer, 2013, pp. 167–176.
- [21] P. M. Musuva, K. W. Getao, and C. K. Chepken, “A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility,” *Computers in Human Behavior*, vol. 94, pp. 154–175, 2019.
- [22] B. Reinheimer, L. Aldag, P. Mayer, M. Mossano, R. Duezguen, B. Loft-house, T. Von Landesberger, and M. Volkamer, “An investigation of phishing awareness and education over time: When and how to best remind users,” in *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, 2020, pp. 259–284.
- [23] D. M. Sarno, J. E. Lewis, C. J. Bohil, and M. B. Neider, “Which phish is on the hook? phishing vulnerability for older versus younger adults,” *Human factors*, vol. 62, no. 5, pp. 704–717, 2020.
- [24] D. M. Sarno and M. B. Neider, “So many phish, so little time: Exploring email task factors and phishing susceptibility,” *Human Factors*, 2021.
- [25] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, “Discriminating stress from cognitive load using a wearable eda device,” *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410–417, 2009.
- [26] T. Van Gog, L. Kester, F. Nievelstein, B. Giesbers, and F. Paas, “Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction,” *Computers in Human Behavior*, vol. 25, no. 2, pp. 325–331, 2009.
- [27] Verizon, “Data breach investigations report 2022,” Verizon, Tech. Rep., 2022.
- [28] A. Vishwanath, B. Harrison, and Y. J. Ng, “Suspicion, cognition, and automaticity model of phishing susceptibility,” *Communication Research*, vol. 45, no. 8, pp. 1146–1166, 2018.
- [29] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao, “Why do people get phished? testing individual differences in phishing vulnerability within an integrated, information processing model,” *Decision Support Systems*, vol. 51, no. 3, pp. 576–586, 2011.
- [30] J. Wang, Y. Li, and H. R. Rao, “Coping responses in phishing detection: an investigation of antecedents and consequences,” *Information Systems Research*, vol. 28, no. 2, pp. 378–396, 2017.
- [31] Q. Wang, S. Yang, M. Liu, Z. Cao, and Q. Ma, “An eye-tracking study of website complexity from cognitive load perspective,” *Decision support systems*, vol. 62, pp. 1–10, 2014.
- [32] R. Wash, “How experts detect phishing scam emails,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–28, 2020.
- [33] R. T. Wright, M. L. Jensen, J. B. Thatcher, M. Dinger, and K. Marett, “Research note—influence techniques in phishing attacks: an examination of vulnerability and resistance,” *Information systems research*, vol. 25, no. 2, pp. 385–400, 2014.
- [34] S. Y. Zheng and I. Becker, “Checking, nudging or scoring? evaluating e-mail user security tools,” in *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*, 2023, pp. 57–76.
- [35] S. Zhuo, R. Biddle, Y. S. Koh, D. Lottridge, and G. Russello, “Sok: Human-centered phishing susceptibility,” *ACM Trans. Priv. Secur.*, dec 2022. [Online]. Available: <https://doi.org/10.1145/3575797>

APPENDIX

STUDY SCENARIO AND INSTRUCTIONS

The following scenario and instructions were provided to our participants. The phishing emails themselves are shown on the next page.

Jacob Smith is an administrative staff member in the CS department, who is organising a CS Collaboration Event. There are speakers from other universities who will present their research in our department. Jacob has been working on managing the speakers' information, their arrival, and other CS event related work, such as booking a room for the event, and setting up catering. However, Jacob is on sick leave now, and we are asking you to cover him and help him set up the event. Your task is to gather event related information from the emails, and put them into three tables.

Your Tasks:

- Read each email carefully and process the emails as if they are sent to you.
- Complete the three tables (the information you need to complete the tables are in the emails).
- If you think an email is important and related to the event, please STAR the email.
- If you think an email is suspicious, please REPORT the email.
- For the last table (Summary table), you will need to add up the corresponding fees from the second table (Transportation table) to calculate the total fee.
- Please inform the instructor when you finish all the emails.

Things to note:

- You should treat each email as if you are responsible for them.
- All the links and attachments are clickable and function as intended.
- Emails are displayed from newest to oldest (oldest emails are at the bottom).
- If an email has attachments, the attachment will be displayed above the reply/forward button.
- Please try NOT to move your head during the study.
- Please try NOT to move your chair at any time.
- Please try NOT to resize the windows.
- There will be TWO 15 minute sessions.
- Please DO NOT discuss this study with other participants because it may bias their behaviour and influence the result.

After reading the scenario and instructions, the participant was invited to ask questions and clarify any aspects of the scenario or instructions that were not clear.

[Important] file shared

Hayley Adaina <hayley235sd@outlook-office.co>
to me

Hayley shared a document


Hayley (hayleyadaina@outlook.com) has invited you to **edit** the following document:


Hey,

This is the budget report you have been asked for the other day, can you have a look and get back to me?

Kind regards,

Hayley

 Budget report



Account on hold emails

nginx user <nginx@vmi398623.contaboserver.net>
to me

You are receiving this message because we have quarantined an email that been sent to [REDACTED] at 09/29/2022 07:04:39 am, with subject line: "IMPORTANT: pending fees need to be processed!".

If you believe the email is legitimate, you can release the email using the link below.

[Release on hold emails](#)

Report needs to be double checked !

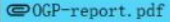
Jay <jay.chapman@tfac.orth>
to me

Hey,

This is the file I mentioned based on our discussion the other day, please have a look and give me some comments.
If you spot any mistakes, please let me know ASAP, I need it by tomorrow.

Cheers,

Jay

 OCP-report.pdf

Product Installation Guide

Quesis Team <quesisteam@kuqpw23.orth.online>
to me

Dear customer,

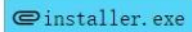
Thank you for ordering our software for non-commercial personal use. You can find your personal Licence Key, download and support information below.

Your Personal License Key (which you need for activating the software) is:
FC66586B-D0F1-4E8B-81B3-BB7AE6525856

Due to some security concern, this Personal license key is valid for only 48 hours. Please make sure you install our software within the next 48 hours use the installer attached in this email.

Kind Regards,

Quesis Team

 installer.exe

The four phishing emails used in the study. a) top left: task-relevant link attack; b) top right: less relevant link attack; c) bottom left: task-relevant attachment attack; d) less-relevant attachment attack.