Impacts and consequences of phishing

Kieran Mead

School of computing

University of Kent

Email km770@kent.ac.uk

**1 Abstract**

This report explores phishing with a specific focus on how phishing attackers can target victims, and whether factors that I will investigate in my project have an influence on potential susceptibility. This has been explored first by extensive literature reviews that I have conducted on papers following a similar theme. To support my conclusions, I conducted a questionnaire that had different scenarios in which participants would respond if they thought it was phishing or not. Most of the papers I have reviewed have a similar study but on a much larger scale. Questionnaires are a good way of gathering what influences someone into thinking an email is genuine or legitimate because that allows you to pick up on people's preconceived ideas of phishing. Here I will list some of the results gathered regarding demographics.

Some summary findings that I found were age and phishing correlation. There is little evidence to support my initial hypothesis based on my findings, but this could be due to a lack of available data variation restricting my ability to conclude this.

Another summary finding I was able to identify was males were broadly speaking less likely to fall for phishing than females. However, I had 13 men to 7 females so there is a bit of a disparity, but this was more reliable than that of age.

Another summary finding was I thought those who use the internet more were less likely to fall for phishing, but this isn’t necessarily true.

The reason in which I did the questionnaire was because the papers that I read had questionnaires in similar ways and looked for similar research questions to that of mine. However, there were ethics limitations that were placed upon the research mainly concerning the distribution method. Originally, I wanted to distribute it to a more varied sample such as family and friends, but this wasn’t permitted by ethics, so I had to compromise and instead make a public social media post that was up for twenty-four hours with the aim of recruiting participants to answer my questionnaire. It wasn’t the most optimal way of finding a varied set of participants but as mentioned I had to get ethics approval, and this was the only way it would be approved.

**2 Introduction**

Phishing is becoming increasingly common, this is supported by one paper I read saying “Phishing attacks are the most common type of cybersecurity breach as stated on government official statistics” (Alkhalil et al., 2021). Attacks are becoming increasingly sophisticated and more plausible which is increasingly causing victims to fall for phishing. Phishing attacks now are designed to target as many people as possible because phishing is so common and is now being run like a business model by phishing attackers. Due to the social engineering aspect of phishing and because we are human any target could potentially be a vulnerable target.

My goal is to increase awareness of the factors that can increase one’s susceptibility to phishing and make people aware so we can prevent people from falling victim to phishing. I created a questionnaire to look at not only some of the demographic factors but also some of the phishing detection reasons, because the questionnaire also asked users what made them think a scenario was phishing or genuine and to give the reasons, this allows me to examine the ones that are picked up on the most as well as those that aren’t. This can educate people about this to reduce their risk in the future. Not only this the questions in the google form also ask basic demographic factors that are not personally identifying, and this includes age, gender, and level of internet use.

This report intends to investigate if there are reasons which can influence whether you fall victim to phishing because of demographic factors like age, gender as well as how often you use the internet. These specific topics are the main focus of my research questions. The reason I selected these as large focuses and set out to answer as my research questions was because many of the existing research papers use these same demographic questions. One such paper is Who Falls for Phish? A demographic analysis of susceptibility (Sheng et al., 2010), so it made sense to study similar questions to that of existing academic research.

Age is an important question because most people would have preconceived views that the older you are the more susceptible you are, but I want to investigate is this always the case? Gender is another interesting research question because unlike age there isn’t really a preconceived answer and from papers I have read, the answer on gender varies, which was interesting so I felt could use some further investigation. And finally, the final demographic question I used was does regular internet use impact susceptibility? This is a very important question because most people would assume that the more you use the internet the less likely you are to be phished, so investigating this would be beneficial as other papers don’t necessarily always look at this as a factor to whether people are more or less likely to be phished.

**2.1.1 Research questions**

The following questions were selected due to papers that I read selecting similar questions. My main research questions are:

* RQ1 How does age impact phishing susceptibility?

This has been addressed in the questionnaire by asking for the age range and then asking them the individual scenarios and whether they would fall for it.

* RQ2 How does gender affect phishing susceptibility?

This has been addressed by the questionnaire by asking for gender and then asking the individual scenarios and whether they would fall for it.

* RQ3 How does regular internet use impact phishing susceptibility?

This has been addressed by the questionnaire by asking for how often you use the internet daily and then asking the scenarios and whether they would fall for it.

**3 Background**

My project is built upon past researchers’ investigations into phishing and what can influence phishing susceptibility. They analyse this in a similar format to my research, which was conducted as mentioned via a questionnaire. RQ1 (which was about age susceptibility) I’ve sought to answer due to many respectable papers in this field looking to whether age correlates with susceptibility. I will explore this further but largely in my data this was not the case. However, this could be due to a lack of variation in my respondent's ages.

**3.1 What is phishing?**

Phishing is a form of semantic attack that exploits the naivety of internet users by deceiving them into divulging sensitive information (Downs, Holbrook, & Cranor, 2007). Many people are aware of phishing but they may not necessarily know what to look for and the techniques deployed by phishing attackers. There are existing phishing mitigation methods that can be used such as URL blacklists (Prakash et al., 2010) and email spam filters that incorporate machine learning as seen in newer papers (Fette, Sadeh, and Tomasic, 2007). However, personally, there is no fool proof way of detecting and preventing phishing but using our judgment alongside sender verification goes a long way toward reducing the phishing chances.

**3.1.1 social engineering in phishing**

Phishing relies on social engineering techniques. Social engineering is defined by Kaspersky (a cybersecurity company) as a manipulation technique that exploits human error to gain information, access, or valuables. (Kaspersky, 2020). Attackers exploit trust amongst people sometimes using the example of business email compromise where attackers prey on human error and pretend to be someone there not i.e. the person’s boss.

**3.2 Phishing detection**

Preventing people from falling victim to phishing attacks has no simple answer but there are multiple factors that can be deployed to detect phishing attempts. As briefly mentioned, one of these which was deployed in a paper I read was spam filters, which are particularly used in organizations and are important to reduce the number of recipients receiving these phishing emails (Fette, Sadeh, and Tomasic, 2007). However, phishing simulations are also important in exposing less experienced employees to what potential risks and features of phishing to look out for. Enabling them to disseminate genuine and not genuine emails more effectively. However, when phishing simulations are carried out it is important to stage them regularly but not too frequently as this can lead to more harm than good (Lain and Kostiainen, 2022).

**3.2.1 Individuals levels of risk acceptance**

Risk acceptance is one of the largest factors that can influence a potential phishing victim. This is due to those who are more risk averse being more hesitant, despite this the impact of social engineering can cloud their judgment into accepting that this person is who they say they are. Risk-taking individuals are even more at risk of phishing exposure due to their less considerate approach to taking risks they are more likely to just do whatever they are asked meaning that someone trying to socially engineer would have a higher chance of success. There is especially an increased threat to specific demographics such as those who perceive the most risk around them, this includes women, and racial and ethnic minorities according to one paper I read named Who falls for phish a demographic analysis of phishing susceptibility and effectiveness of interventions? (Sheng et al., 2010a).

Also, how participants respond to simulations demonstrates that despite training being important because training isn’t a real-world scenario, they are more likely to be more relaxed and not necessarily consider the real-world impact if this has happened to them, which doesn’t benefit those undergoing the simulation to put what they have learned into practice. As the NCSC says, “simulations can erode trust, Employees should instead create positive cyber security culture” (NCSC, 2018). Therefore, leading to a better working environment for employees and educating them on phishing risks and not punishing employees.

**3.2.2 Impact of trusting emails**

A huge influencer on phishing susceptibility not only relates to risk but is largely linked to the role of trust, and the role of trust can cause you to fall victim to phishing. Not only this many victims of phishing attacks admitted one of the main consequences was a reduced level of trust in the future. This is supported in one of the papers I read in which the writer surveyed 155 people, many said a consequence of them falling victim to phishing was reduced trust (Carroll, Adejobi and Montasari, 2022). It is important to acknowledge that falling victim and reducing trust isn’t necessarily a good thing, many people think that reduced trust of websites and emails is a good thing but this isn’t always the case because what happens when you see a legitimate email or website but you question it, and this can lead to false detection rates. Also, trust plays a key role in the most severe phishing campaigns due to the nature of how trust works, an example of this could be that your boss asks you to do something via email, and because it is your boss you do and reply to them, but did you ever consider that might not be your boss and could be someone impersonating your boss for phishing and social engineering purposes. And if you are more likely just to trust the email author you are less likely to scan the email for typical phishing cues and this has an impact on your trust with platforms in the future (Carroll, Adejobi and Montasari, 2022).

**3.2.3 summary of background**

The main factors that I believe impact this is gender and age but perhaps some of the other demographic questions that people don’t consider like risk acceptance levels and phishing detection ability, so can you pick up on subtle email and website phishing attempts as well as trust levels? All these factors can impact an individual’s ability to fall victim to a phishing attack. Some of these demographic questions were raised in a google form questionnaire I created and distributed via social media. I acknowledge due to the distribution method much of the respondents were of largely male and of similar age range (18-25). However, I can still use my data to try and draw conclusions based on this.

**4 Methodology**

**4.1 overview**

Figure 1 below depicts the key stages of the project.

A diagram of a process

Description automatically generated

Figure 1 Methodology flow diagram showing process of research.

**4.1.1 paper review**

Began by reading academic papers some of which have informed my conclusions, and some are illustrated in the background.

**4.1.2.1 questionnaire design**

I designed my questionnaire based on the paper review questionnaires and aimed to ask similar questions see 4.1.2.2 and 4.1.2.3 for some examples from the questionnaire. The reason I mainly focused on age, gender and technology use was because the papers I read identified these as questions to explore.  **More example scenarios can be found in the corpus.**

**4.1.2.2 Questionnaire scenario example**

To illustrate how the questionnaire worked I am going to write one of the scenarios that was asked in the questionnaire here to give an indication of what they were asked to analyse. Responses were recorded anonymously. An example is below.

**Hi John,**

**Thanks for your recent purchase of the iphone 14 we are preparing it now and it will be with you in the next 4-6 business days. If you have further concerns about the product, please do not hesitate to contact us at the email address above again. Once dispatched you can use this link to track it** [**https://www.tracking-at-phonewarehouse.co.uk**](https://www.tracking-at-phonewarehouse.co.uk)

**Kind regards**

**Phone warehouse team.**

**4.1.2.3 Questionnaire contextual scenario example**

Some scenarios were given additional context to make sure they fully understood the scenario before answering. For one of my scenarios, it was regarding financial and banking,

So it was key I told them that the scenario were a customer of Halifax bank, so I reduced responses of people asking how do we know he banks with Halifax and restrict responses to phishing based?

**In this scenario you receive an email this is an email from Halifax Bank which is who John has an account with.**

**4.1.3 Questionnaire Hypothesis**

* Does Age correlate with phishing susceptibility such as are older people more likely to fall for phishing vs younger people?
* Does gender correlate with phishing susceptibility i.e. are male vs female more or less likely?
* Does internet usage correlate with phishing susceptibility?

**4.1.4 Questionnaire distribution**

The questionnaire was distributed via social media, specifically an Instagram story which contained a link to the questionnaire and that is how responses were collected. I acknowledge this can lead to a lack of varied age range and gender due to the profile of those who are on social media.

Ethics approval was required prior to distribution and hence the method of distribution was social media.

**4.1.5 response analysis**

the responses were exported into a spreadsheet containing all the information, and this allowed me to make charts as seen in the results section of this report.

**5 results**

Here I will convey the main results concerning my research questions. I received 20 responses to the questionnaire.

**5.1 Age and correlation of phishing susceptibility**

The below bar chart shows Age of all participants who answered.

The bar chart shows the age ranges of participants who answered the questionnaire.

For word clouds see the results spreadsheet for full context contained within the corpus, if you would like to see the full reasons given.

A graph with blue squares

Description automatically generated

Figure 2 Bar chart showing Age

**Scenario 1**

The below chart shows the yes and no responses for whether an email was genuine or not, where yes is genuine, and no is phishing. Scenario 1 shows that in the 18-25 range all participants believed scenario 1 was phishing. A similar trend continued with the 35-50 range. However, the only 50+ participant believed it was phishing and wasn’t genuine.A graph with red and green squares

Description automatically generated

Figure 3 shows the age correlation of whether a scenario was genuine or not for scenario 1

Below is the word cloud of why participants thought it was or wasn’t phishing.

The word cloud below is one of the main reasons scenario 1 was deemed phishing or wasn’t deemed to be phishing by participants. The reason the word bad is so high up is because this was referencing bad grammar in scenario 1, and also looks was up there for looks like phishing. The reason email is up so high is because lots of people said the email looked legitimate although this scenario was meant to be an SMS. Order is high up because the user wrote that it was an order he placed. **Full reasons can be found in the results spreadsheet located in corpus.**

A close-up of words

Description automatically generated

Figure 4 scenario 1 word cloud reasons for why or not phishing.

**Scenario 2**

The below chart in scenario 2 shows the yes-no responses where yes means genuine and no means not genuine. As we can see in scenario 2 all age respondents believed this one was phishing which was what I predicted due to being quite an obvious phishing scenario. However, what is good here is all respondents correctly identified the scenario was phishing despite any age factors, showing in this example certain practice like sharing of username and password is universally accepted as phishing despite age.

A graph with red rectangles and green text

Description automatically generated

Figure 5 shows age correlation of whether a scenario was genuine or not for scenario 2

Below is the word cloud of why participants thought it was or wasn’t phishing.

This below word cloud is like what I predicted regarding scenario 2 because scenario 2 was such an obvious phishing scenario that asked for a username and password. It was understandable that not only all respondents to the questionnaire identified this as malicious, but also identified it largely from the same reason for it asking for the username and password of employees. The reason password is so large is because it asked for username and password and this was what users picked up on the most. There was nobody who deemed this no phishing hence there is only the red word cloud. **Full reasons can be found in the results spreadsheet located in corpus.**

A close-up of words

Description automatically generated

Figure 6 scenario 2-word cloud for reasons why or why not its phishing

**Scenario 3**

The below chart in scenario 3 shows all the genuine or not responses, where genuine is yes and not genuine or phishing is no. As we can see 4 people in the 18-25 age range identified this as genuine compared to 11 who said that it was phishing. Looking to the 35-50 age range, we also have 1 person who believed it was genuine compared to 3 who said it was phishing. In the 50, + range they thought it was all genuine. I was a bit surprised about this one because there were account details provided in this scenario. So, I am surprised all ages apart from the 50+ range were divided.

A graph of a number of people

Description automatically generated

Figure 7 shows age correlation of whether a scenario was genuine or not for scenario 3

**below is the word cloud of why participants thought it was or wasn’t phishing.**

The below word cloud is largely what I expected to have the bank number and account number as high up on the list also call because many respondents said they would call their bank if they received an email from them. However, perhaps people didn’t read the scenario context just above the body of text because some people said that in the answers he might not have Halifax bank. However, the scenario says he does bank with Halifax. The reason bank shows up so large on the word cloud is because lots of people said it had specific details such as bank details and bank numbers. Those who thought it was phishing picked up on don’t click links hence don’t and click and link is so large. **Full reasons can be found in the results spreadsheet located in corpus.**

A close-up of a number

Description automatically generated

Figure 8 shows word cloud for scenario 3 reasons why or why isn't phishing

**Scenario 4**

the below chart shows for scenario 4 all the genuine or not genuine responses where yes is genuine, and no isn’t. The 18-25 range thought the email in this scenario was phishing this trend continues with 35-50. However, 50+ said it was genuine and not phishing which is correct for this specific scenario it surprised me because I expected it to be the other way around the younger people would have gotten this correct, but this shows the importance of false positive results when it comes to phishing. However, this could have been caused like the other scenario in which they hadn’t read the context above the scenario because it does say that John placed this order himself so some of the responses questioning whether he placed the order are incorrect.

A graph of a bar chart

Description automatically generated with medium confidence

Figure 9 shows age correlation for whether a scenario was genuine or not for scenario 4

**Below is the word cloud associated with reasons given as to why it was or wasn’t phishing.**

The word cloud is what I was expecting considering it was an Amazon delivery email and the

main words it detects are understandable considering it is about a delivery. The main conclusion from this word cloud was people didn’t think it was phishing because it contained a delivery number or an order number. However, those who said it was phishing said it was phishing because it didn’t include a delivery address. The reason order is so large is due to people largely picking up on the order number and specific information. However, the reason delivery address shows up on the not genuine side is due to it not mentioning delivery address. Only one respondent provided a response for phishing in this scenario. **Full reasons can be found in the results spreadsheet located in corpus.**

A close-up of words

Description automatically generated

Figure 10 Showing word cloud for scenario why or why it was or wasn’t phishing.

**5.2 results for Gender correlation of phishing**

A graph with a red and green rectangle

Description automatically generated

Figure 11 showing gender split in questionnaire.

**Scenario 1**

the below chart shows the split between male and female and whether they deemed scenario 1 genuine or not genuine. 3 Males thought it was genuine 10 Males thought it was not genuine and 3 females thought scenario 1 was genuine compared to 4 Females who thought it wasn’t genuine.

**A graph of a bar chart

Description automatically generated with medium confidence**

Figure 12 shows a correlation between gender and whether scenario 1 was genuine or not

**Scenario 2**

The below chart shows the split between male and female and whether they deemed scenario 2 genuine or not genuine. All males and females believed this was not genuine which I expected considering what the scenario was.

**A graph with red squares

Description automatically generated**

Figure 13 shows the response to scenario 2 whether it was genuine or not by gender.

**Scenario 3**

The below chart shows the split between males and females and whether they deemed it genuine or not for males vs females for scenario 3. Females have 3 who said it was genuine and 4 who said it wasn’t genuine. Compared to men we have 3 who said it was genuine and 10 who said it wasn’t genuine.

**A graph with red and green squares

Description automatically generated**

Figure 14 showing response to scenario 3 whether it was genuine or not by gender.

**Scenario 4**

The below chart shows the split between male and female and whether they deemed scenario 4 genuine or not. As we can see 6 females said scenario 4 was genuine whereas 1 said it wasn’t genuine and for men, all respondents said it was genuine.

**A graph with a red and green bar chart

Description automatically generated**

Figure 15 showing response to scenario 4 whether it was genuine or not by gender.

**5.3 results for internet use**

The 2nd demographic question I asked was internet usage daily this would allow me to answer one of my main questions (RQ3), which was does regular internet use reduce the risk of falling victim to phishing. The below pie chart shows the % of participants who said they used the internet anywhere ranging from 2-5+ hours. in the results below 5 hours are in the green, 4 hours is in the orange, 3 hours is in the red and 2 hours is in the blue, but blue is only 1 respondent.

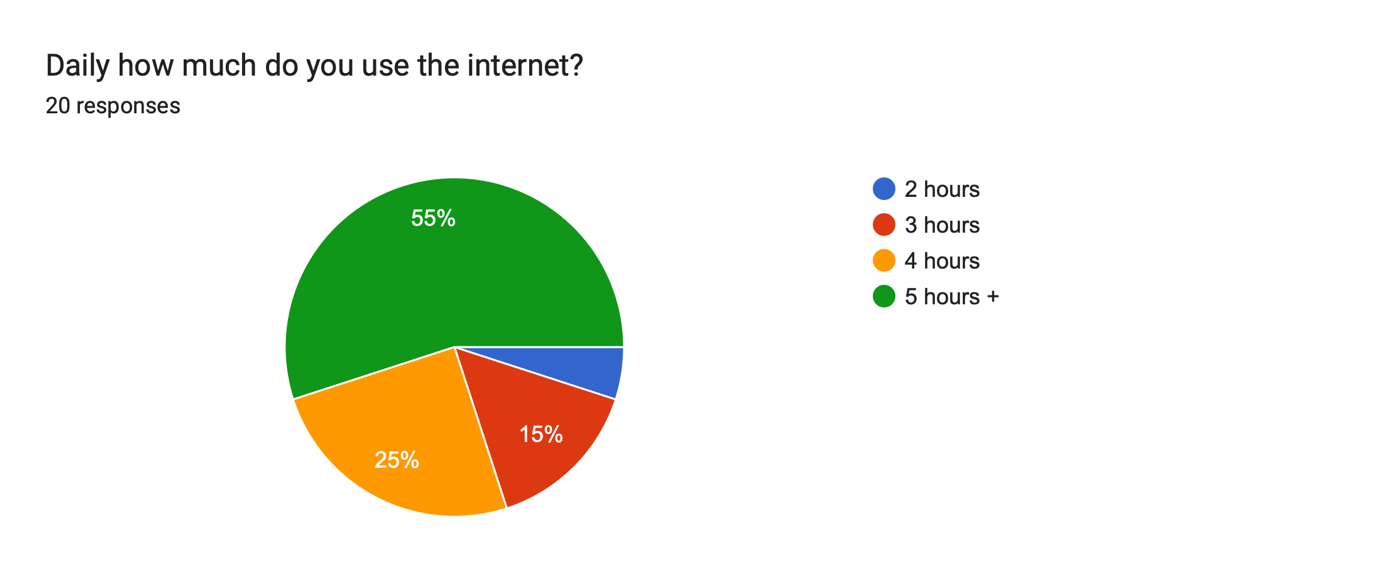


Figure 16 Shows the Daily internet usage of participants measured in hours.

**Scenario 1**

the below chart shows the % of questionnaire respondents who deemed scenario 1 to be genuine or not and their corresponding time spent on the internet each day. As we can see 3 people said they use the internet for 3 hours a day 2 said it was genuine and 1 said it wasn’t. 5+ hours was the most common daily usage result and therefore out of 11 people, 9 said scenario 1 was not genuine and 2 said it was genuine. For the 4-hour usage result, there were 5 respondents 2 of which said it was genuine and 3 of which said it wasn’t. Finally, the only 2-hour respondent said that it wasn’t genuine.

A graph of a bar

Description automatically generated with medium confidence

Figure 17 Showing Internet usage and phishing perception in Scenario 1

**Scenario 2**

The below chart shows the % of questionnaire respondents who deemed scenario 2 to be genuine or not and their corresponding time spent on the internet each day. For all respondents, I expected this result due to the obvious nature of the phishing scenario, but the result was all respondents no matter how long they spent on the internet all said it was phishing.

A graph with red squares

Description automatically generated

Figure 18 Shows internet usage and phishing perception in scenario 2.

**Scenario 3**

The below chart shows the % of questionnaire respondents who deemed scenario 3 to be genuine or not and their corresponding time spent on the internet each day. For the 3-hour respondents, 1 person believed scenario 3 was genuine compared to 2 who said it wasn’t. Of the 5 hours+ category 4 people thought it was genuine compared to 7 who didn’t think it was. For 4-hour respondents, 1 person said it was genuine compared to 4 who didn’t. And finally, the only 2-hour response was not genuine.

A graph with red and green squares

Description automatically generated

Figure 19 shows internet usage and phishing perception in scenario 3

**Scenario 4**

The below chart shows the percentage of questionnaire respondents that deemed scenario 4 to be genuine or not and their corresponding time spent on the internet. As we can see for three hours, four hours, and two hours all responses were that the email was genuine. However, one respondent in five hours+ said it wasn’t genuine. The one respondent that was an outlier in the data said it wasn’t genuine and claimed this as the reason why “Doesn’t have a delivery address”.

A green and red graph

Description automatically generated

Figure 20 Shows internet usage phishing perception in scenario 4

**6 conclusions**

**6.1 Conclusion for age**

Initially, I believed that younger individuals would be less likely to fall for phishing attempts. However, when examining how participants perceived phishing messages in Scenario 1, I found that those aged 18-25 were more inclined to view them as genuine. Although this difference wasn't significant, it indicates the need for further investigation. To better understand this, it is important to help illustrate that there is variation within age ranges which demonstrates this further.

In Scenarios 2-4, where participants assessed the genuineness of emails, I didn't observe a clear relationship between age and perception. This lack of correlation made it difficult to confirm my initial hypothesis. Despite anticipating that younger individuals would be less susceptible to phishing across these scenarios, the data showed a similar level of perception across age groups. Therefore, age doesn't seem to play a significant role in how individuals perceive email genuineness based on my findings. As mentioned in the first paragraph, some of the scenarios had variations within age groups making it hard to make a conclusive result.

It's important to acknowledge that my dataset has its limitations. By including a wider range of age groups and more participants, I could have conducted a more thorough analysis to explore the connection between age and susceptibility to phishing attempts. However, I felt it was worth the time to try and explore this as many of the research papers in this area investigated this. The factor played upon age and phishing susceptibility, and if there was a correlation. One paper I read (Baiting the Hook factors impacting susceptibility to phishing attacks) found that Age and a detection score correlation didn’t exist. (Iuga, Nurse, and Erola, 2016). This is like that of my findings for age that there wasn’t a correlation between age and being able to detect whether it is genuine or not. This is important to use to support my findings because this paper had a questionnaire participant count of 382 and more varied ages were surveyed. (Iuga, Nurse and Erola, 2016).

**6.2 Conclusions for Gender**

Initially, I believed that men would be less likely to fall for phishing them women. Figure 11 shows the distribution, and this quite clearly shows that 65% of the participants were male and 35% were female any data we analyse we must first acknowledge that there is a bit of a skew towards there being more men than women.

In scenario 1 figure 12 shows that more males (specifically ten) believed that the scenario was not genuine compared to three males who thought it was genuine. When comparing this to females as mentioned we must acknowledge there were only seven females. However, four females believed it wasn’t genuine compared to three who thought it was genuine. This shows here that men were more sceptical and less likely to fall for this phishing attempt.

In scenario 2, figure 13 shows males and females unanimously agreed that scenario 2 was phishing which I expected due to the nature of the scenario asking for username and password this is universally agreed as bad practice to share so I expected this result for this scenario which is encouraging. This shows a shared level of phishing recognition across genders in certain scenarios. However, most phishing scenarios would be more believable than these nowadays.

In scenario 3 figure 14 shows a difference in responses between males and females. While females were evenly split in their perception with three marking it as genuine, and four marking it as not genuine. However, Males largely perceived this as not genuine with a similar trend to scenario 1 with ten of the male participants marking scenario 3 as not genuine and three marking it as genuine. Overall, this scenario implies that there is a divergence with males being more cautious once again. However, I could have improved my study if I included indicators aimed at identifying false positives, because some of these respondents no doubt would have labelled phishing emails as genuine and non-genuine emails as phishing. Therefore, in scenario 3 men were less likely to fall for phishing in this attempt.

In scenario 4 figure 15 shows a notable difference in male and female responses. Most females 6 out of 7 believed scenario 4 was genuine whereas all male respondents concluded that scenario 4 was genuine. This shows that most of both males and females believed this was genuine and only one out of the twenty responses said it wasn’t genuine. However, due to all men saying this was genuine, we can conclude that there isn’t too much of a difference in this scenario, but due to the only variance being in the female category, we can say females are more likely to perceive the phishing as not genuine. However, scenario 4 was intended to be legitimate so the majority got this correct but there was one respondent who identified this as not genuine in the female category.

Overall, the trend amongst gender and correlation to phishing susceptibility shows given my data in scenarios 1 and 3, more males perceived phishing attempts as not genuine compared to females. This shows a trend that men are more sceptical to trust these phishing scenarios. However, in scenario 4 females had one respondent who said it was not genuine which in this instance indicates a false positive for phishing because this scenario was intended to be genuine.

Therefore, this illustrates that there is a variability amongst males and females but the reasons for this difference aren’t clear in my questionnaire as I did not investigate this, but perhaps the content in the scenarios can influence the outcome of the questionnaire. Overall, men were more sceptical than women but also, there were more male respondents than females so given the data I had my hypothesis was correct but as mentioned my study was limited in scale.

**6.3 Conclusions for Internet Usage**

Initially, I believed that more regular users would be less likely to fall for phishing. Starting with figure 16, this shows the variation of internet usage among respondents, the majority declared they used the internet for five hours+ then four hours was the second largest then three hours then two.

In scenario 1 figure 17 we can see that most participants are using the internet for three, four, five hours+. However, despite many respondents using the internet for longer periods the majority believed this was not genuine. However, the trend indicates that they believed this one was not genuine, despite this some outliers did believe it was genuine. In this scenario, the respondents who indicated they only used two hours also said that they thought it wasn’t genuine. So, therefore, in scenario 1 my hypothesis was disproved.

In scenario 2 figure 18 we can see that the data here was unanimous and this was expected due to the scenario being obvious to all users that this was a case of phishing. All respondents no matter the internet use level said that it was not genuine. This also somewhat disproves my hypothesis because the person who only used it for two hours was able to disseminate this but as mentioned this scenario was obvious to most participants.

In scenario 3 figure 19 we can see that the data here was somewhat more of a split amongst three, four , five hours with one out of three in the three hours saying it was genuine and one out of 5 in the four hours and in the five hours + there was a split. Most users said scenario 3 was not genuine which was somewhat unexpected I intended this scenario to be genuine, but most participants believed it was not. Therefore, in this scenario, the hypothesis that those who use it more are less likely is once again not true because the majority of three, four, five hours + voted in the same way. And the person who used the internet for two hours also voted the same as the majority.

In scenario 4 figure 20 the data shows here that almost all of the respondents got this one correct. This was indeed another genuine scenario, so this result was somewhat expected. However, once again my hypothesis is disproven because the only person who thought it was not genuine was in five hours+ and the person in the lowest category was two hours got this correct.

Overall, we can conclude internet usage doesn’t impact susceptibility as I thought in my hypothesis based on the data I gathered from the questionnaire. Despite varying internet usage levels among participants it is clear to see from the results that even in the higher levels of usage like four , five hours they weren’t always correct in their perception and also there was a divide among usage groups i.e. some in the five hours thought it was genuine and some didn’t which surprised me as I assumed those that use the internet more would know more of what to look for. I would expect the same results amongst those in the same use level that were five+ hours. However, the data here is the most varied of all my 3 research questions and RQ3 has the most varied data responses which enabled me to investigate this research question most effectively of all 3 questions.

**Overall conclusion**

Overall, although the data gathered doesn’t necessarily prove my original hypothesis based on my research questions, I still feel it was a valuable task to do based on research papers that indicated trends within the data. However, although it doesn’t prove my entire hypothesis we can still read into the data and see that demographic factors do play a role due to the variation amongst those identifying in the same category as their peers selecting differing opinions based on the same scenarios. A paper I read (Baiting the hook factors impacting susceptibility to phishing attacks) supports this “Studies discovered that gender, age, and exposure to materials have some impact “(Iuga, Nurse, and Erola, 2016).

To further improve my study if I had more time, it would have been beneficial to expand my sample size and gather a more distributed cohort, but due to constraints on project time and ethics approval granted based on distribution via social media, it restricted my sample largely to 18-25-year-olds and the majority of these were male and due to ethics restrictions the youngest age range I could do was 18+.

**Acknowledgements**

Thanks to my Supervisor Ozgur Kafali for supervising the project.

Thanks to the author of the papers, I have used for my research.

Thanks to all the participants of my questionnaire.

# Bibliography

Abroshan, H., Devos, J., Poels, G. and Laermans, E. (2021). COVID-19 and Phishing: effects of human emotions, behaviour, and demographics on the success of phishing attempts during the pandemic. *IEEE Access*, 9, pp.1–1. doi:https://doi.org/10.1109/access.2021.3109091.

Alkhalil, Z., Hewage, C., Nawaf, L. and Khan, I. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*, [online] 3(1). doi: https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060/full

Iuga, C., Nurse, J.R.C. and Erola, A. (2016). Baiting the hook: factors impacting susceptibility to phishing attacks. *Human-centric Computing and Information Sciences*, [online] 6(1). doi:https://doi.org/10.1186/s13673-016-0065-2.

Carroll, F., Adejobi, J.A. and Montasari, R. (2022). How Good Are We at Detecting a Phishing Attack? Investigating the Evolving Phishing Attack Email and Why It Continues to Successfully Deceive Society. *SN Computer Science*, 3(2). doi: https://link.springer.com/article/10.1007/s42979-022-01069-1.

Downs, J.S., Holbrook, M. and Cranor, L.F. (2007). Behavioral response to phishing risk. *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit on - eCrime ’07*. doi:https://doi.org/10.1145/1299015.1299019.

IBM (2023). *Cost of a data breach 2023*. [online] IBM. Available at: https://www.ibm.com/reports/data-breach.

Verizon (2023). *2023 Data Breach Investigations Report*. [online] Verizon Business. Available at: https://www.verizon.com/business/resources/reports/dbir/.

Wang, J., Herath, T., Chen, R., Vishwanath, A. and Rao, H.R. (2012). Research Article Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email. *IEEE Transactions on Professional Communication*, 55(4), pp.345–362. doi:https://doi.org/10.1109/tpc.2012.2208392.

NCSC (2018). *Phishing attacks: defending your organisation*. [online] www.ncsc.gov.uk. Available at: https://www.ncsc.gov.uk/guidance/phishing#.

Prakash, P., Kumar, M., Kompella, R.R. and Gupta, M. (2010). PhishNet: Predictive Blacklisting to Detect Phishing Attacks. *2010 Proceedings IEEE INFOCOM*. doi:https://doi.org/10.1109/infcom.2010.5462216.

Kaspersky (2020). *What is Social Engineering?* [online] usa.kaspersky.com. Available at: https://usa.kaspersky.com/resource-center/definitions/what-is-social-engineering [Accessed 11 Mar. 2024].

Fette, I., Sadeh, N. and Tomasic, A. (2007). Learning to detect phishing emails. *Proceedings of the 16th international conference on World Wide Web - WWW ’07*. doi:https://doi.org/10.1145/1242572.1242660.

Lain, D. and Kostiainen, K. (2022). *Phishing in Organizations: Findings from a Large-Scale and Long-Term Study*. [online] Available at: <https://arxiv.org/pdf/2112.07498.pdf>.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L.F. and Downs, J. (2010). Who falls for phish? *Proceedings of the 28th international conference on Human factors in computing systems - CHI ’10*. doi:https://doi.org/10.1145/1753326.1753383.

Moody, G.D., Galletta, D.F. and Dunn, B.K. (2017). Which phish get caught? An exploratory study of individuals′ susceptibility to phishing. *European Journal of Information Systems*, 26(6), pp.564–584. doi:https://doi.org/10.1057/s41303-017-0058-x.

P. K. Yeng, B. Yang, M. A. Fauzi and P. Nimbe, "A framework for exploring incentive methods towards reducing phishing susceptibility in Healthcare: Based on a review and In-the-wild-field study approach," 2023 Intelligent Methods, Systems, and Applications (IMSA), Giza, Egypt, 2023, pp. 228-234, doi: 10.1109/IMSA58542.2023.10217499.

Tembe, R., Zielinska, O., Liu, Y., Hong, K.W., Murphy-Hill, E., Mayhorn, C. and Ge, X. (2014). Phishing in international waters. *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security - HotSoS ’14*. doi:https://doi.org/10.1145/2600176.2600178.

Sumner, A. and Yuan, X. (2019). Mitigating Phishing Attacks. *Proceedings of the 2019 ACM Southeast Conference on ZZZ - ACM SE ’19*. doi:https://doi.org/10.1145/3299815.3314437.

Abroshan, H., Devos, J., Poels, G. and Laermans, E. (2018). Phishing Attacks Root Causes. *Lecture Notes in Computer Science*, pp.187–202. doi:https://doi.org/10.1007/978-3-319-76687-4\_13.