An educational experiment in discovering spear phishing attacks

Sebastian Floderus
Linus Rosenholm

Faculty of Computing, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden
This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfilment of the requirements for the degree of Bachelor of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

The authors declare that they are the sole authors of this thesis and that they have not used any sources other than those listed in the bibliography and identified as references. They further declare that they have not submitted this thesis at any other institution to obtain a degree.

Contact Information:
Author(s):
Sebastian Floderus
email: sefl16@student.bth.se

Linus Rosenholm
email: lirb16@student.bth.se

University advisor:
Senior lecturer, Martin Boldt
Department of Computer Science

Faculty of Computing
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden
Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57
Abstract

Background: Spear phishing attacks uses social engineering targeting a specific person to steal credential information or infect the users computer with malware. It is often done through emails and it can be very hard to spot the difference between a legitimate email and a scam email. Cybercrime is a growing problem and there is many ways to inform and educate individuals on the subject.

Objectives: This study intends to perform an experiment to see if an educational-support tool can be used to better identify phishing emails. Furthermore see if there is a difference in susceptibility between students from different university programs.

Methods: A qualitative research study was used to get the necessary understanding how to properly develop a phishing educational tool. A Pretest-Posttest experiment is done to see if there is an improvement in result between an experimental group that received education and the control group that did not.

Results: The result shows an overall higher score for the technical program compared to the non-technical. Comparing the pretest with the posttest shows an increase in score for the non-technical program and a decrease in score for the technical program. Furthermore 58% of the non-technical students who started the test did not complete it.

Conclusions: There is a noticeable difference in susceptibility between the programs for detecting scam emails for students. However further research is needed in order to explore to what extent the education process had an impact.

Keywords: Phishing, Spear-phishing, Experiment, Security awareness
Acknowledgments

We want to say thank you Martin Boldt for guiding us through our work and helping us with valuable feedback.
# Contents

Abstract i  
Acknowledgments iii  

1 Introduction 1  
1.1 Background ........................................................................... 1  
1.2 Problem motivation ................................................................. 2  
1.3 Aim of study .......................................................................... 3  
1.4 Research questions ............................................................... 3  
1.5 Scope ..................................................................................... 3  

2 Theory 5  
2.1 Phishing background ............................................................. 5  
2.1.1 Social engineering ............................................................. 5  
2.1.2 Spear phishing .................................................................. 5  
2.1.3 Clone phishing .................................................................. 6  
2.1.4 Website phishing ............................................................... 6  
2.2 Why phishing works ............................................................. 6  
2.2.1 Using emotions as bait ...................................................... 7  

3 Related work 9  

4 Method 13  
4.1 Qualitative and quantitative research ....................................... 13  
4.1.1 Qualitative research .......................................................... 13  
4.1.2 Quantitative research ......................................................... 13  
4.1.3 Mixed method ................................................................. 13  
4.2 Design of experiment ............................................................. 14  
4.2.1 Educational tool ............................................................... 14  
4.2.2 Survey .............................................................................. 15  
4.2.3 Ethical aspects ................................................................. 16  
4.2.4 Statistical method ............................................................. 17  

5 Results 19  

6 Analysis and discussion 21  
6.1 Research questions ............................................................... 23  

7 Conclusions and future work 25
References

A Supplemental information

A.1 Education process .................................................. 31
A.2 Email tests ........................................................... 32
Chapter 1

Introduction

1.1 Background

Phishing is a type of criminal attack using social engineering to steal credential information from an individual or a company. Attackers use spoofed emails claiming to be a legitimate business, the message will re-direct the user to a counterfeit website with the objective to steal personal information such as username and password or infect the users PC with malicious software [1]. However phishing attacks are massively spammed out to everyone and are therefore easier to spot as spam emails. Meanwhile a spear phishing attack is more sophisticated and designed to target a specific individual, group or company. Since the attacker limit the targets, it is easier to get a well crafted message with personalized information like your name and/or your job title that makes the email look more believable [2].

Phishing is not a new subject, the first publicly known phishing attack took place sometime around 1995 and targeted America Online (AOL), one of the biggest providers of internet access at the time. Hackers used email posing as AOL-employees to trick users into giving up information such as passwords or credit card numbers. Since phishing was a new concept many fell for it despite warning signs such as a abundance of grammatical errors [3, 4]. Even though phishing attacks have been around for 25 years it is still as relevant as ever. According to Symantec’s Internet Security Threat Report of 2018, 71% of all infection vectors (how an attacker gets access to a victim’s network) was done through spear phishing emails [5].

In 2015 the Ponemon Institute released a report on the cost of phishing attacks and the value of training employees. The report was based on a survey made on 377 IT organizations in the USA. 39% of the participants had over 1000 employees. The report concluded that the annual cost of phishing attacks for a average sized company was $3.77 million. Out of this number $1.8 million or $188.4 per employee could have been saved by performing anti phishing training [6].

There is a big difference in today’s society compared to two or even one decade ago. Internet today is a tool used daily for all sort of things, checking emails, social media, gaming, marketing, banking etc. Also more and more companies have digitized their infrastructure, however the knowledge of IT-Security is lacking, it is predicted that there will be 3.5 millions unfilled cybersecurity jobs by the year 2021 [7]. It is also predicted as the population of internet users is rapidly growing that the
cost of cybercrime will have an annual cost of $6 trillion by the year 2021 compared to $3 trillion in 2015 [8].

Spear phishing attacks are a very common tool of attacks and it can be used for both smaller attacks but also for attacks on a country’s infrastructure. An example of this can be the attack against the Ukrainian power grid in 2015 [9]. In this case spear phishing was used among a series of other attacks eg. BlackBerry 3 malware (a trojan used for DDOS-attacks, espionage, and destruction of information), theft of business network credentials and social engineering. The spear phish was successfully used to gain access to the business networks of the oblenergos (a term used to describe an energy company) to eventually get a foothold in these companies. The result of the attack led to several outages which affected approximately 225,000 customers to lose power for several hours across various areas.

Another spear phishing attack that occurred back in 2009 was an attack called Phish Phry. This attack was carried out in the United States and was targeted towards bank account holders. The FBI back then described it as the one of the largest cybercrime investigation ever carried out in the United States [10]. The investigation resulted with about one hundred individuals charged with crimes including computer fraud, conspiracy to commit bank fraud, money laundering, and aggravated identity theft. By stealing hundreds of account holders financial information the attackers used it to transfer about 1.5 million dollars over to fake accounts that they controlled.

There are many different types of strategies to protect users from falling prey to phishing attacks but generally speaking they fall under three major categories. This paper will focus on the 3rd option, anti phishing education.

1. **Automatic removal**: This tactic removes the threat without interacting with the user. E.g. email service providers can detect spam email and automatically remove them.

2. **Warning**: Many tools exist designed with the purpose to warn users about phishing attacks. E.g. Browser toolbars that warns a user when visiting a potential phishing website.

3. **Training**: There are many different approaches to training users to better identify phishing. E.g. A tool to test users ability to identify phishing emails.

### 1.2 Problem motivation

It does not matter how strong the firewall is, if one use cryptography and the most expensive Anti-virus software, the weakest link will always be the human factor of any security system. Normal people is at risk of being exposed to identify theft and credit card fraud [11]. Meanwhile as previously shown in the section background, companies can stand to risk millions in revenue. Cybercrime is already a big problem and will only keep growing in the future. Spear phishing is a very simple but effective type of attack. Some people will naturally be less susceptible to phishing attacks than other. It is all about knowing what to look for, something anyone can learn.
1.3  Aim of study

The aim of this research is to study different types of spear phishing emails, further investigate why it is such an effective tool for hackers. In addition perform an education experiment to see if some groups of people is at higher risk to become victims than others, moreover evaluate if there is a increase in result after the education.

1.4  Research questions

To be able to reach the goal of this project the following research questions will be answered:

1. What different types of phishing emails exist based on the email content and to what extent do they share common identifiers?

2. To what extent can an educational-support tool aid users in identifying spear-phishing attack emails?

3. Is there a difference in susceptibility to spear phishing attacks for students at non-technical educational university programs compared to students enrolled in technical programs?

1.5  Scope

This research strongly focus on spear phishing, however other types of phishing attacks will be discussed to get a basic understanding of the phenomenon and better understand what is typical indicators for identifying these types of attacks.

The study will be limited to semantic attacks, i.e attacks that focuses on how we as humans interact with systems and interprets information [12], e.g. modifying information in an email in such a way that it looks correct but in reality it is misleading. With that said it will not focus on malware based phishing attacks.

The target group for the experiment will be students attending Blekinge Institute of Technology (BTH), the study was limited to this since we as the research group have good connections to resources at BTH and also so that not to much work is taken on for the amount of time of the research.
Chapter 2

Theory

2.1 Phishing background

This section will give an overview of different types of phishing techniques, what is the difference and what characteristics do they share? Furthermore terms and topics related to phishing will also be discussed, this is to get the reader to understand the basic concept of phishing. There are many types of phishing attacks, phishing done through social media, SMS (Short Message Service) or over the phone, malware-based phishing and the list keeps going. However the ones discussed in this section are most relevant to this research and spear phishing most often combines elements from both clone phishing and website phishing. Elements such as cloning a legit email and attach a link to a counterfeit website [13].

2.1.1 Social engineering

Social engineering is a technique where a person targets another individual through manipulation to retrieve valuable information. It is used for various reasons whether it is to get a raise at your job or just to get ahead in line at the supermarket. Social Engineering is used by everyday people and it is not necessary something that is used only for criminal activities [14]. However, unfortunately, it is used by criminals to bypass security either physically or by making someone give up sensitive information. These types of attacks focuses on exploiting humans instead of technical systems. People are weak towards manipulation, things such as respect for authorities, curiosity and compassion are a few examples of emotions that hackers can use against their target. [15].

2.1.2 Spear phishing

The subject of spear phishing have already been touched on in the introduction and as previously mentioned this phish have been specifically targeted. At a minimum the attacker knows the following parameters of the target: first name, last name and email address. From this information it is probably easy to find ones social media account, information from the persons workplace or anything else that you have signed up for on the internet. Depending on the person and the amount of digital footprints that is left behind a lot of information can be found, hobbies, family members, job titles etc.. [16]. Social engineering is also a tool that can be used when compiling a spear phishing email.
One strength of spear phishing emails is that the victim can get the feeling that there have been a previous relation and contact with the sender which will make it more trustworthy. From the mentioned information a well crafted message can be made that will be very hard to identify as phishing.

### 2.1.3 Clone phishing

Clone phishing is when the attacker creates a replica of a real message that has already been sent to a user. The email is sent from a address that is spoofed to look like the original sender, e.g. Instead of support@github.com, the address might look like support@githud.com. Everything looks the same, the text, the attached images, however the link or attached files in the email have been swapped out with malicious content. The attacker can motivate the duplicates of messages saying that it is an updated version or some sort of error in the previous message [17]. It can also be an email that is warning the receiver about an issue with their account eg. a google account. They then urge the victim to follow the link below which leads to a replica of the original site where he or she will “update” the account information. Clone phishing is not necessarily but can often be a subcategory of spear phishing.

### 2.1.4 Website phishing

Website phishing is mostly used to target one individual rather than a whole organization. it is fairly easy to counterfeit a replica of legitimate website. The goal can be to steal personal or financial information, e.g. Copy the design of a banks website and steal the log-in information from the victim. This type of attack is often combined with email phishing [13].

### 2.2 Why phishing works

There are many reasons behind why people fall for phishing attacks, the two biggest ones are lack of knowledge and visual deception.

1. **Lack of knowledge:** Many users of the internet lack the basic understanding of how it works, how email, applications and the operating systems works and interact with each other. For example not everyone understand the syntax of domain names and will not be able to differ a legitimate URL versus a spoofed URL. E.g. the domain member-paypal.com is not connected to the domain paypal.com, meanwhile member.paypal.com is a subdomain to paypal.com.

![https://www.paypal.com](https://www.paypal.com)

Figure 2.1: SSL

Another example is the little padlock next to the URL as presented in Figure 2.1, many users lack the knowledge that this locked padlock indicates that
2.2. Why phishing works

SSL is activated and a secure encrypted connection is established between the browser and the web server [18].

2. Visual deception: Even if users don’t fall under the category "lack of knowledge" they can still fall prey to phishers. Visual deception is a technique used to trick users by mimicking the content of an email or a website, copying images, text and fonts. A common method is using an image of a real link to a website but in reality the image is a hyperlink to another website [18].

2.2.1 Using emotions as bait

There are many factors going into the decision making process of humans. Phishers try to target victims and make them forget about logic thinking by appealing to the emotions of the target [19].

![PayPal Image](Image)

**Figure 2.2:** Targeting the greed of users

Figure 2.2 illustrates a simple example of phishers trying to appeal to the greed of people. It is $45 free! All you have to do is click on the link. This example is not a very strong phishing email, it can be spammed out to thousands of people and only a small percent will click the link. Figure 2.3 shows a more well crafted real example of an spear phishing email that circulated the web, that a lot more people fell pray to[16].

![Spear Phishing Image](Image)

**Figure 2.3:** A real example of a spear phishing email (The contact information have been altered)
The message looks very believable, even a person that is educated in the subject might become a victim. Breaking down Figure 2.3, we can see that:

1. The real district court logo of the U.S is used.
2. The phisher has done research and uses the full name and address of the target.
3. The email used is "@uscourts.com", this looks like the real deal except that the real US court uses ".gov" instead of ".com".
4. The attacker is using fear and respect for authorities as well as stressing that the matter is urgent.

[16]
A lot of studies have been made on the subject of phishing with different methods and results. In this chapter a short summary of relevant studies will be presented, what kind of studies have been done before and what the result was.

In 2018 Kaspar Jüristo conducted a simulated email phishing experiment [20] which was made to explore how to create consolidated guidelines that companies could easily implement and use as a standard to better detect phishing emails. Furthermore at the same time take in to consideration how people react to it so that they do not make the subjects feel at unease. They also wanted to find out the correlation between the phishing email difficulty level and the click through rate. This was done by having two test groups “K” and “L” where both groups did two different tests that have different difficulty levels. What they found out is that in the easier tests that had more obvious signs like misspelling, 11% fell for the phishing mail whereas for the more complex ones, 23% were baited. Some of the subjects also agreed to take an additional test where they would be specifically targeted other known as whaling. By doing this the bait accuracy reached 100% but the subjects integrity were more at jeopardy which was debated.

An email phishing survey was done in 2018 [21], a research about phishing knowledge amongst employees at Blekingesjukhuset (The Hospital of Blekinge) in Sweden Karlskrona. The survey tried to find out the risk of falling victim to a ransomware attack through phishing emails targeting employees. What is concluded from the results of the article is that the subjects did not know at all what to look for in an email to properly detect legit and non legit emails. One of the flaws with this study was that there were a low number of participants which will affect the accuracy of the result.

In a study from 2009, School of Phish: A Real-World Evaluation of Anti-Phishing Training [22]. A study with the main purpose to educate users to better make decisions when identifying phish and none phishing emails. One of the challenges they experienced was that in general people are not motivated to learn about security and they feel like it is a secondary task. To counter this a tool and a game was developed called PhishGuru and Anti-Phishing Phil. PhishGuru is a tool used for direct educational response for when the user opened their normal emails where as the Anti-Phishing Phil is a game for educational purpose. What they found out was that tools like this can effectively teach people how to avoid phishing attacks and
they also believe that it can be used for other security purposes. However humans knowledge about phishing should be combined with automated detection systems as a first line of defence against phishing emails.

Back in 2015 a quantitative study was made to measure performance in detecting of phishing emails. The tests was divided in two groups, test one and two that had ten emails in each part. The tests had five phishing and five legitimate emails in each part and the two different parts were done as similar as possible e.g. To have the same amount of emails from company X. The participants would then do one of the tests, after do the training part and then continue with the opposite test. The study resulted in that even though the tests were somewhat effective it was not enough to make sure that the participant could avoid phishing emails all together [23]. In the follow-up study they are going to include more tests and they will also change so that there would be a direct response after each decision in every email.

One approach to stop phishing emails is to use different types of machine learning techniques. Machine learning is used to detect regular spam emails but it can also be used to detect specific emails like phishing emails. In a study from 2007 they used machine learning to make specializing filters [24] to detect phishing emails based on content attributes. It would make decisions based on either the internal information in an email which would be the actual text and how it is formulated or it will check external sources, links to websites and more. It was found out that it was possible to detect phishing emails with high accuracy using features that are more directly applicable to phishing emails than to regular spam. They believed that this can be because regular spams do not have to misrepresent their identity unlike a phish email.

A proposition in one article [25] is to use training interventions for phishing website detection and that they believe end-user training is a key component to mitigate phishing exploitation. The training intervention used in this study will let the end-user try to detect phish on their own and if they make a mistake there will be helpful information provided immediately in an attempt to improve the end-users detection competence. The main purpose of this study was to compare the efficiency of this training approach with outmoded methods such as sending anti-phish emails that provide tips on how to detect phish emails. In addition, there was also a control group. What they concluded from this study was a significant positive effect using this new method to help users properly judging legitimate and Phishing websites.

A seemingly common concept of training in attempts to improve people’s phishing detection skill is to implement embedded training. It is hypothesized in a study that this type of training, where there is an immediate response when the end-user is making a bad judgmental decision. To further explore the efficiency of embedded training a large-scale experiment was conducted to analyse end-users reaction towards spear phishing put together with the embedded training. Contrary to what earlier studies suggested it was more difficult to make this type of training efficient enough in the corporate environment it was set in. Their result suggested that immediate feedback is not sufficient enough to reduce click ratings. Based on the amount of time the end-users spent on every part they could assume that they did not read
all the training [26].
Chapter 4

Method

This chapter will go through the different methods used over the duration of the project, a motivation for each method of choice will be presented. Likewise a description for the following parts: the data gathering process, the structure of the experiment, as well as how the collected data will be analyzed.

4.1 Qualitative and quantitative research

Generally speaking research methods can be categorized under either a qualitative study or a quantitative study and in some cases a mixed method. Both methods will be discussed in addition to a motivation which will be used for this study.

4.1.1 Qualitative research

The objective of performing an qualitative research is to gather a broad range of information to get a deeper understanding of a phenomenon. It is not meant to test a hypothesis. The information is often gathered through interviews or reviewing documents. The sample group is typically small [27].

4.1.2 Quantitative research

The objective of performing an quantitative research is finding a relationship between independent variables and other dependent outcome variables. The information is gathered through an experiment or a survey. Then the result is analyzed to test the strength of the hypothesis given. Quantitative research focuses heavy on numbers and logic to determine a relation between a theory and empiricism. The sample group is typically large [28].

4.1.3 Mixed method

Both types of method has their strengths and weaknesses, however in some cases the two methods can be combined to reinforce the weakness of each method. This will generate more information on the subject and strengthen the credibility of the analysis [29]. Since the goal of this research is to analyze and find relationships between different target groups a quantitative method will be used in the form of an experiment. However a qualitative method in the form of gathering information prior to the analysis and the development of the educational-phishing tool is necessary.
4.2 Design of experiment

Many different types of experimental researches exist, among other: Post-test only, Pretest-Posttest and Solomon Four Group design [30]. A randomized block two group pretest-posttest design will be used for this study which was the design that felt most suitable for the purpose of this research. It is a common method that can be used for testing the effect of a variable (the education), it ensures that the experiment has a high level of internal validity since the pretest makes sure that the groups are equivalent. Once the test is complete the posttest result can be compared between the groups to see to what extent the education helped [31]. Let's elaborate what a randomized block two group pretest-posttest means:

1. **Randomized Block**: The participants are first divided into blocks based on their field of study before being randomly assigned between the experimental group and the control group.

2. **Two group**: The experimental group will receive education on the subject of phishing in hopes of improving the result. The control group will receive alternative information not relevant to identifying spear phishing emails.

3. **Pretest-posttest**: Measurement is done before and after participants come in contact with the educational process.

As illustrated in Figure 4.1, the structure of the experiment is the same for both groups, this is the case to try to eliminate as many external factors that might change the outcome of the results. For example the "Alternative Education" will have no real significant purpose beyond keeping the tests equal for both groups, thus the amount of text and the complexity is designed to be similar for both "educations".

4.2.1 Educational tool

A educational tool has been developed in the form of a web application, as shown in Figure 4.1. The application will streamline the process of gathering data in the form of multiple surveys, as well as categorizing the data based on the students field of study.
The decision was made to use a web application tool instead of simply using images for identifying phishing emails. Because with an interactive web tool it is possible to get a more real life experience, as presented in Figure 4.2. More information about the email is accessible such as sender, date and subject. Also it is possible to hover over a link and see the actual URL.

![Figure 4.2: Example from one of the tests in the application.](image)

When first accessing the application participants are greeted with a short information section about the experiment as well as the reasoning behind the project. The user can then input information about their field of study. From here students are randomly divided into an experimental group and a control group. Both groups will perform the same survey based on phishing, the result will be collected. Next, the experimental group will perform a quick educational training on the subject of email phishing, meanwhile the control group will receive alternative information with fun facts about phishing (see Appendix A). In the last step both groups will receive a new survey with new questions that will also be collected (see Appendix B for all the test emails). Furthermore the participants will see their result of the test. The data is collected in to a MySQL database so that the result later can be analyzed.

### 4.2.2 Survey

As previously discussed a questionnaire method will be used. Each individual will do two separate surveys were they are evaluating in total 12 test emails such as the one presented in Figure 4.2. Half of the emails are legitimate and the other half is phish. Likewise the participant will have the option to mark the emails as "Legitimate" or "Phish" and for every right decision the participant will gain one point. Table 4.1 lists all the emails and their characteristics based on the information:

1. Personalized info: The email targeted a specific person that was mentioned by name in the content text.

2. Fake URL: The real URL was hidden behind either an image or a hyper-link.
3. Grammatical errors: The content either had spelling errors or that the text was poorly formulated.

<table>
<thead>
<tr>
<th>Email</th>
<th>Type</th>
<th>Personalized info</th>
<th>Fake URL</th>
<th>Grammatical errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Github</td>
<td>Legitimate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 PayPal</td>
<td>Phish</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>1.3 Netflix</td>
<td>Legitimate</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4 Google docs</td>
<td>Phish</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>1.5 Amazon</td>
<td>Phish</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>1.6 Google</td>
<td>Legitimate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 UPS</td>
<td>Phish</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2.2 Ericsson</td>
<td>Legitimate</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3 Zalando</td>
<td>Legitimate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4 Citi bank</td>
<td>Phish</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2.5 Google</td>
<td>Legitimate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.6 Apple</td>
<td>Phish</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 4.1: Describing the email properties.

The sample group will be limited to students that are enrolled to a program at Blekinge Institute of Technology. Advice has been offered to keep the research on a smaller scale and also to achieve the goal of this research. Therefore focus will solely be on students from the IT-Security department compared to the nursing department. As well as the option "Other" that represent all other participating programs.

4.2.3 Ethical aspects

When designing a phishing experiment it is important to remember that the study includes real people with real information. There is a lot of phishing experiments that are based on the idea of deceiving people, the participants are not aware that an experiment is in motion. This can be very important to simulate a real world problem and thereby get data that is closer to a real scenario. However there are some ethical complications, there are laws and regulations that must be followed when storing information about participants. Phishing may cause stress or psychological damage depending on the content of the email, also people are unpredictable and may act differently when responding to a email [32]. E.g a person receives an email demanding the individual to pay a fine or appear and testify in court. One of the participants may choose to contact their lawyer another one tries to pay the fine.

When using the experimental application for this study the user is in a virtual email application, no personal information is saved and all the participant are fully aware that an experiment is on going. The test is voluntary and there is no consequences of not participating or not finishing the test.
4.2.4 Statistical method

The outcome of the experiment will be analyzed with statistical tests. The simplest variant of this is to calculate the gain score:

\[ gain = posttest - pretest [33] \]

of every individual participant, then add them together and calculate the mean score. The experimental group will be compared to the control group, furthermore the result of the nursing program will be compared to the result of the IT-security program.

From the result of the gain score and mean values a more sophisticated test can be made, to see if there is a difference between the groups. A paired sample t-test is a valid method that is commonly used to test if there is statistical evidence that the mean difference between two measurements is significant different from zero [34]. Another method to use is Cohen’s d which is appropriate when two groups have the same size in addition to a similar standard deviation. Cohen’s d is a common method used to calculate the increase of effect between two different groups by taking the mean difference and dividing them by the combined standard deviation [35]. Furthermore the data will be analyzed to see if there is a time difference and a difference in the way participants report emails between the pretest and posttest.
Chapter 5

Results

The results were collected over a period of one week, in this time 40 participants completed the whole experiment out of 72. Among these participants, 19 came from the nursing program and 21 from the IT-Security program. The maximum score for each part of the test is six points, combining the pretest and the posttest this adds up to 12 points. Figure 5.1 shows an overview of the mean value for the participants in every group for the pre-test and the post-test, difference shows the increase or decrease in score.

Figure 5.2 shows the result of every group for the pretest and the posttest. The data that is shown is maximum, minimum, lower quartile(25%), upper quartile(75%) and the median. The top and bottom of the vertical line represents the minimum and maximum, the lower and top part of the box represents the quartile and the dark blue line is the median.

Figure 5.1: The overall result of the experiment for every group.
Figure 5.2: The result presented in a box plot.

<table>
<thead>
<tr>
<th>Program</th>
<th>Overall</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse Experimental</td>
<td>3.9</td>
<td>3.7</td>
<td>4.15</td>
</tr>
<tr>
<td>Nurse Control</td>
<td>4.6</td>
<td>4.65</td>
<td>4.55</td>
</tr>
</tbody>
</table>

Table 5.1: Showing the mean values of nurse compared to IT-security

Among all the participants three people managed to get a perfect score of 12, all of these students came from the IT-security program. Furthermore there was a large amount of students who began the test without completing it. Out of the 72 partakers, 45 came from the nursing program and 27 from the IT-security program. 26 of the nurses or 58% dropped out of the research before completing the experiment. Meanwhile 6 of the security students or 22% did not complete the test.

Looking at the different groups a more detailed result can be displayed when measuring the different answers and time. The following is the average number of times phish was clicked for part 1 compared to part 2 and the average time difference between the pre-test and post-test. If the time is negative it means that the time have improved by X seconds. The percentage that is shown is the accuracy of which the groups could detect the phishing emails.

<table>
<thead>
<tr>
<th>Group</th>
<th>Phish Part 1</th>
<th>Phish Part 2</th>
<th>Time Difference(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse - Experimental</td>
<td>2.5 (57%)</td>
<td>2.6 (67%)</td>
<td>-58</td>
</tr>
<tr>
<td>Nurse - Control</td>
<td>2.4 (61%)</td>
<td>3.7 (88%)</td>
<td>-47</td>
</tr>
<tr>
<td>IT-Sec - Experimental</td>
<td>3.1 (91%)</td>
<td>2.9 (79%)</td>
<td>+6</td>
</tr>
<tr>
<td>IT-Sec - Control</td>
<td>2.9 (70%)</td>
<td>3.9 (91%)</td>
<td>+6</td>
</tr>
</tbody>
</table>

Table 5.2: Showing the average time difference and the average amount of times phish was reported for the pretest and the posttest.
Based on the result gathered from the study we can see a clear improvement for the nurses of the experimental group +0.8 points compared to the control group, the control group still had an improvement but only by 0.1 points which is rather small. Furthermore the IT-security students had the reversed result where the experimental group had a decrease by -0.6 points, compared to the control group that had an improvement by 0.4 points.

The result of the nurses was the most clear difference, the experimental group had an increase in value of 13.3%, this can be an indicator that the education process had a correlation with the increase in score. The control group that had no education only had an increase of 1.6%. However here the pretest also had a higher value compared to the experimental group, 4.3 against 3.1, this could also be a contributor to the low increase in score. Hence with more participants the control group and the experimental group should logically have about the same start value for the pretest.

A paired sample t-test was done for each program, one for the nursing program and one for the IT-Security program presented in table 6.1 and table 6.2. It is a test designed to see if the effect of a variable(score difference) have a significant difference from 0 between the experimental group and the control group. Group statistics show some describing statistics like the mean value and standard deviation for each group. The more interesting facts can be found in table 6.2, Independent sample T-test. Levene’s test, that is a part of the T-test was used in order to see if the variance within the group is the same or not. Consequently if the Levene’s significant threshold is below 0.05 means that equal variance can not be assumed. However if the significant threshold is above 0.05, equal variance can be assumed [36].

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean difference</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse - Experimental</td>
<td>9</td>
<td>0.777</td>
<td>1.9220</td>
</tr>
<tr>
<td>Nurse - Control</td>
<td>10</td>
<td>0.100</td>
<td>0.9944</td>
</tr>
<tr>
<td>IT-Sec - Experimental</td>
<td>10</td>
<td>-0.600</td>
<td>1.5055</td>
</tr>
<tr>
<td>IT-Sec - Control</td>
<td>11</td>
<td>0.454</td>
<td>1.3684</td>
</tr>
</tbody>
</table>

Table 6.1: The group statistics from the T-test.
Table 6.2: The result of the independent sample t-test

<table>
<thead>
<tr>
<th>Group</th>
<th>Levene’s sig</th>
<th>Sig(2-tailed)</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse - Equal variances assumed</td>
<td>0.044</td>
<td>0.430</td>
<td>-0.6777</td>
</tr>
<tr>
<td>Nurse - Equal variances not assumed</td>
<td>0.361</td>
<td>0.109</td>
<td>1.0545</td>
</tr>
<tr>
<td>IT-Sec - Equal variances assumed</td>
<td>0.923</td>
<td>0.111</td>
<td>1.0545</td>
</tr>
<tr>
<td>IT-Sec - Equal variances not assumed</td>
<td>0.111</td>
<td>0.109</td>
<td>1.0545</td>
</tr>
</tbody>
</table>

The Levene’s test for the nursing students, 0.044 is below the threshold, therefore equal variances cannot be assumed. In table 6.2, looking at the row Nurse - Equal variance not assumed and the column Sig(2-tailed). Since the value 0.361 is bigger than 0.05, as a result it can not be said with certainty that there is a significant difference between the groups according to the t-test. Meanwhile the levene’s test for the security students, 0.923 is above the threshold, therefore equal variance can be assumed. However the value for Sig(2-tailed), 0.109 is also bigger than 0.05, hence there is no significant difference between the groups for the security additions either.

Another test called Cohen’s d was made to measure how large the effect of the education process was, by using the means and standard deviation of two groups:

\[
Cohen's \, d = \frac{M_1 - M_2}{\text{Std} \, \text{dev}_{\text{pooled}}} [37]
\]

Nurse Cohen’s d = 0.4424, this indicates that the effect size is small/medium.

IT-security Cohen’s d = 0.7328, this indicates that the effect size is medium/large.

The outcome of the two tests concluded that for both programs the difference found in the random selection cannot not with certainty be found for a greater population within 95% accuracy. It is still a possibility but the outcome could be a coincidence. This could be an effect of the groups being to small since it was only approximately 10 persons per group. Furthermore the average score of the pretest between the control group and the experimental group was quite different, Nurse: Experimental(3.1), Control(4.3) and IT-Security: Experimental(5.1), Control(4.2). For a bigger test group these values should be closer to each other. However the result of the T-test does not strengthen the hypothesis that the education had any impact. In addition higher educated students from the IT-Security program could possibly all ready known some of the information in the education meaning it had a smaller effect. Furthermore the standard deviation between the experimental group and control group for the nurses was very large, as previously mentioned the standard deviation need to be similar to each other to be considered a definite Cohen's d value.

In table 5.2 detailed group statistics, the result was left much up for speculation. Both experimental groups had no significant difference in number of reported phish emails from part one and part two in the test. The control groups however, both showed a clear increase in number of clicks for phish in the second part. There is
no evidence why this occurred but it is possible that the information provided about phish for the control group may have made the participants more aware of the phenomenon phishing. As a result this may have affected their ability to believe they spot more emails of the type phish.

As mentioned in the related work section, it can be very challenging to get regular people interested in IT security [22] [26]. 58% of the nurses who started the test did not complete it, compared to the IT-Security where 22% did not finish the test. Furthermore the time difference illustrates a big difference between the pre-test and post-test for the two nurses groups, an increase of speed in the second part with close to a minute quicker. Time difference and dropout rate can have a correlation, meaning the longer the test went on the lack of interest grew and thereby participants wanted to finish the test faster. Another alternative is that the nurses got more familiar with the test and gained an understanding what to look for quicker. However the IT-Security students had no significant change in time difference.

6.1 Research questions

**RQ1:** What different types of phishing emails exist based on the email content and to what extent do they share common identifiers?

This question was designed to help us develop a better phishing education tool and is foremost answered in the chapter theory.

**RQ2:** To what extent can an educational-support tool aid users in identifying spear-phishing attack emails?

The education process seems to have had a small impact but further testing is required to get a definite answer.

**RQ3:** Is there a difference in susceptibility to spear phishing attacks for students at non-technical educational university programs compared to students enrolled in technical programs?

There is a clear difference in susceptibility between students from the nursing program compared to the IT-security program, however if more university programs were included in the study a different result may have occurred.
Chapter 7

Conclusions and future work

Spear phishing attacks is a big issue and has been around since 1995 and the attacks are increasing in number every year and does not seem to be decreasing anytime soon. The consequences of falling prey can be significant, for individuals but even more for companies. As previously mentioned in the problem motivation of the introduction section. Normal people is at risk of being exposed to identify theft and credit card fraud [11]. Meanwhile companies is at even higher risk, company secrets leaking out could lead to millions of loss in revenue.

The result gathered from the test illustrates that there is a clear difference in susceptibility to phishing attacks between the nursing program and the IT-Security program, the security students had an overall higher score. We can see an increase in score for the experimental group for nursing students and a decrease for the security students, furthermore none of the t-test showed a significant difference. Therefore it can not be said with certainty that the education process had an impact or if it was a coincidence. Further testing on a larger scale is necessary to get a more definite result.

With the high dropout rate for nursing student in combination with the increase of speed on the posttest the conclusion can be made that it is most likely harder to get non-technical students interested in security issues, when they most likely are the ones needing it more.

The research questions formulated for this study have been overviewed, analyzed and answered. Because of the few number of participant in this study the t-tests can become more precise in future studies if it is done on a larger scale. A similar study can be made with more participants and more university programs. It is possible that without human interaction it is difficult to get people involved and interested in the subject if they are only forced to read. If the pedagogical reading part was longer and more detailed, it is possible that a larger percentage would lose interest after a while. Therefore, another way of educating is encouraged.
References


Appendix A

Supplemental information

A.1 Education process

Facts about phish.

How to identify a phish education.
A.2 Email tests

Question 1.1, legitimate.

Question 1.2, Phish.
A.2. Email tests

Question 1.3, Legitimate.

Question 1.4, Phish.
Question 1.5, Phish.
A.2. Email tests

Question 1.6, Legitimate.

Question 2.1, Phish.
Appendix A. Supplemental information

Question 2.2, Legitimate.

Question 2.3, Legitimate.
Question 2.4, Phish.

Question 2.5, Legitimate.
Dear Dave,

We have received your iTunes account is used for fraud. Your account will be suspended until you confirm that you are the original user account. To confirm that you are the original user of this account:

[Click here]

/iTunes security team

Question 2.6, Phish.