TALLINN UNIVERSITY OF TECHNOLOGY

Faculty of Information Technology

Department of Computer Science

ITC70LT

Panagiotis Marzelas 130548

A SOCIAL MEDIA HONEYPOT METHOD TO DETECT SPEARPHISHING

Master thesis

Olaf Manuel Maennel

PhD

Professor

TALLINN 2015

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Marzelas Panagiotis

05.01.2016

Abstract

In this document we propose a method to detect spearphishing with the use of social media honeypots in order to protect employees from disclosing information about other employees or assets of an organization. The attackers gather information from multiple sources for a victim and by applying psychological techniques they try to exploit the weakest link in the cyber security chain, the human vulnerabilities that exist in each personality. Our goal is to reverse that process by exploiting the human element of the attackers and to deceive them by creating false social network profiles for fake employees. We believe that if there is no direct way to contact a high profile employee the attackers will target a person from his social or working environment to gain his trust and use him as a stepping stone. Our work is based on two hypotheses, first that the spearphishers study the psychological profile of the target prior to attack in order to apply the correct psychological techniques and second is that if we create a fake person which carries the personality characteristics the attackers look for, then they will choose to approach that person.

The people who interact with the fake social network accounts are deceived and therefore we discuss the legal and ethical considerations that occur. The European Directives about e-privacy and data protection provide the ground to work with social honeypots without abusing the collected personal data. The sharing of the honeypot data between the companies and public sector could create an interesting partnership against spearphishing that could lead to a safer online environment.

This thesis is written in English and is 78 pages long, including 11 chapters, 8 figures and 8 tables.

Annotatsioon

Suunatud andmepüügi avastamine sotsiaalmeedia meepoti meetodiga.

Lõputöö on kirjutatud Inglise keeles ning sisaldab teksti 78 leheküljel, 11 peatükki, 8 joonist, 8 tabelit.

Table of abbreviations and terms

The table of abbreviations and terms must consist of all new and ambiguous terms. For example the abbreviation PC can be used as Personal computer or Program counter. Abbreviations and terms are presented in two columns where the left column consists of term or abbreviation and the right column gives explanation of it. Foreign words are presented in italic. The following example presents a correctly formatted table of abbreviations and terms.

HPE	High Profile Employee
PC	Personal Computer
PA	Personal Assistant
DoD	Department of Defense
U.S.	United States
E.U.	European Union
NSA	National Security Agency

Table of contents

List of figures

List of images enlists only images presented in the main part of the thesis; images in appendixes are excluded.

List of tables

List of tables enlists only tables presented in the main part of the thesis; tables in appendixes are excluded.

1. Introduction

Nowadays it is commonly accepted that the weakest link in information systems is the computer user [1]. There are examples showing that even professionals in information security might be victims of a cyber attack based on social engineering like phishing and spearphishing. CIA's Director recent email account hack [2] or the research project with the code name "Getting in bed with Robin Sage" [3] support that argument. That is why many types of attacks are focusing on affecting the emotional state of the user and make him take decisions that normally would not take [4]. To have better results the attackers use information from many different sources but it seems that the easiest are the social networks in which people freely share their personal information. Even though there have been plenty of reports highlighting and warning the use of social media features in spearphishing, it seems that it does not exist much literature which studies this phenomenon [5]. That was our motive to work on this project and create a social media honeypot and a method that would link spearphishing with the published information on social media.

All companies, organizations, Ministries etc use a set of protection systems, from a simple antivirus to a sophisticated defense system, against outside and inside threats. One such system is the honeypot. Honeypots are systems used in an organization to lure the attacker to act against them by giving him[1] the impression that he is trying to hack the organization's main system. That way the attacker is kept busy while the security team of the organization can take action against malicious activity and study the techniques used by the potential hacker. There are two main categories of honeypots, production and research honeypots [6]. Production honeypots are the ones used by most companies and corporations and they add value to the organization because they protect in a direct way while research honeypots add little value but they are used to study the techniques of the BlackHat[2] community. In this project we will propose a method to detect phishing and spearphishing with the use of social media honeypots.

Phishing and Spearphishing are types of Social Engineering attacks and they are

[1] We use him or he instead of him/her, he/she etc
[2] BlackHat community, http://blackhat.community/

conducted mainly through email. An attacker with a social engineering is trying to deceive a targeted person in order to obtain information about him or someone else. Phishing and spearphishing attacks are mainly used to extract financial gain or intellectual property from the targeted person. They might also be used to obtain security related information about an organization that will be used in an attack against that organization.

Phishing emails are sent to hundreds or thousands of persons addressing to the recipient in an abstract way such as “Dear client” claiming to be from some authority like a banking institute. They make use of a convincing story in order to lure the victim to open an attachment, click on a link or install some kind of malicious software. On the other side spearphishing can be defined as a targeted phishing attack. The attackers know about the victim a big amount of personal information that have been harvested from his social environment, an insider or social networks and they are addressing to the victim with his name like “Dear John”.

Phishing and spearphishing attacks are deployed in 4 phases, Target, Lure, Hook and Catch [7] and it is suggested to be studied for attribution purposes as 4 separate eCrime cases. During Target phase an attacker is searching information about the person in target and it is the most difficult stage to have signs of. The Lure phase is the convincing story in the email that will lure the victim to obey to the instructions, for example, to click on the provided link. That link in the email when is clicked starts the Hook phase and the user will be redirected in a webpage that looks similar to the one the user recognizes as legitimate and he will submit the information asked. After the victim submits his personal information during Hook phase the Catch phase starts which is the procedure the attacker will obtain this information without being caught.

The reason such attacks are gaining more ground every year is because they have big financial gain and it is difficult for the law enforcement to find the intruders and bring them to justice. Mikko Hypponen of F-Secure referred on eCrime as a social problem and as such is very difficult to be tackled. He said[3] “people with skills without the opportunities will have to make a living. If they can’t make a legal living they will make an illegal living.”

[3] Mikko Hypponnen at Estonian IT College, http://www.itcollege.ee/blog/2015/10/13/avalik_loeng/ , 13-10-2015.

Studies have shown [4] that the attackers are using specific psychological techniques to affect the emotional state of a future victim and obtain the information they are after. Such techniques involve persuasion, impersonation, ingratiation, conformity and friendliness. Pressures of conformity, compliance and obedience cause people to change their behaviors [4]. The attackers make use of persuasive techniques like authority, urgency, politeness and they use convincing stories about email account verification, personal information upgrade etc.

The aforementioned techniques aim to exploit existing vulnerabilities in the personality of a computer user. The personality can be analyzed according to the theory of "five traits of personality" or Big-5 [8],[9],[10]. According to Big-5 the personality has 5 main traits, Neuroticism, Extroversion, and Openness to Experience, Agreeableness and Conscientiousness. Every of these traits have its own characteristics and by applying the correct psychological techniques they can be exploited. For example, people high in neuroticism mainly express negative emotions, persons high in extroversion tend to expose their personal life, people high in Openness to Experience are open to new and unusual ideas, persons high in Agreeableness tend to believe that the other people have good intentions and finally persons high in Conscientiousness are showing high responsibility and they are focused on their duties.

During the project we present information on how a person behaves in Facebook according to the personality trait he belongs into [9]. Research papers [10], [11] have found the most popular words people are using according to their personality trait. Additionally [12] and [13] present a correlation between the expressions used by potential insiders, their personality trait and the YouTube and Twitter activity.

As it is mentioned during the Target phase there are normally no indications that someone is looking information on the organization or an employee. In this project the scenario will protect a high profile employee of a company from a spearphishing attack. In order to create a one way road from the attacker to the high profile employee through the honeypot, the high profile employee must create first a circle of trust which he will be connected to in social networks, receive from and send emails to. Research proposes the use of decoy elements for such operations [7]. Our proposed honeypot will be used to collect information during an attacker's Target phase and possibly predict a future attack on the company.

An administrator will create fake employees as decoy elements. The fake employees will have everything a real employee has, an office telephone number, job position, job background, corporate email, a CV, Facebook and LinkedIn accounts and possibly company's website references in order to create results in the search machines when someone will look information on them.

The administrator will combine information gathered from literature and will manage the fake employees' Facebook profiles accordingly, in order to give the impression of a vulnerable person or an insider. Facebook and LinkedIn profiles will be most of the time the communication of the fake employees with the outside world. Social networks will give the opportunity to the attacker to study the online behavior of the fake employees and infer his personality. The goal is that the attacker will be deceived and will decide to contact or approach the fake employee. If that happens, it will give a strong argument that the intruders study the psychological profile of a target and they act accordingly. At the same time, an email will arrive in the organization's mail server for the fake employee. That will alert the administrator for a possible phishing or spearphishing attack. Then, the administrator will collect the available information in the email and all incoming emails from that address (or the emails containing the phishing email's reply_to address) will be dropped. The organization might share the results with the national CERT or other organizations for better protection of private and public sector employees.

The use of honeypots in general raise some questions because if the honeypot is exploited it might be used as a stepping stone for an attack against multiple targets. At the same time honeypots cannot be widely known that they exist because that would cancel their purpose. It is possible that even real employees of an organization would be deceived or communicate with the honeypot and disclose personal information. Honeypots also gather information for the employees' online behavior which raises also some questions whether an organization has the right to study the personality aspects of its employees. The general discussed issues with the honeypots are entrapment, privacy and liability. The issues with our honeypot are online impersonation and deception. Privacy matters in U.S. are covered by the U.S. Privacy Law[4] and in E.U. by the e-

[4] Privacy Act of 1974, "The Privacy Act of 1974, 5 U.S.C. § 552a, establishes a code of fair information practices that governs the collection, maintenance, use, and dissemination of information about

Privacy [14] and Data Protection Directives [15]. The law provides the ground to create a security mechanism with respect to the personal data of the employees.

The management of the fake social media profile was time consuming because the administrator will have to interact with the new friends, gain their trust and create communication channels that will give him in return "likes", comments, tags and an online activity that is very important in order to form the desired personality. During the experiment we build a method to create fake social media profiles as social media honeypots but we did not manage to detect spearphishing attacks. We were contacted online by email, Facebook and LinkedIn by people with suspicious intensions and these cases will be analyzed in the results section. We would need more time to verify that these results occurred because of our work with the fake social network profiles. An opportunity to deploy the proposed social honeypot in a real company with real threats, we assume, that would provide more clear results.

Public and private sector should invest in social media honeypots because by taking proactive measures move the defense perimeter further from the target. We constructed social media profiles that reflect certain personality traits that could be used as honeypots for spearphishers. We did not verify that the spearphishers study the psychological profile of the victim prior to an attack but we believe that it is a matter of time to start doing it. That is because, as we already know, spearphishers study their victims prior to an attack and there are online tools that automate the procedure of analyzing a personality with information taken from social media networks. For that reason, the next step for a spearphisher to have better results is to link the two procedures.

---

individuals that is maintained in systems of records by federal agencies. A system of records is a group of records under the control of an agency from which information is retrieved by the name of the individual or by some identifier assigned to the individual.", http://www.justice.gov/opcl/privacy-act-1974.

2. Hypothesis

Our experiment will be based on the fact that according to literature [16], [4] social engineers manage to exploit human vulnerabilities "via psychologically constructed communications" which show that the attackers have great knowledge of cyber-psychology. Additionally, the spearphishers as social engineers collect large amount of information about their victims from different sources, including social media networks, before conducting the spearphishing attack. They use psychological techniques to affect the emotional state of the target, in order to make him take decisions that normally he would not take [4]. Thus, our first hypothesis is that the spearphishers study the psychological profile of a computer user prior to attack through social networks in order to apply the correct psychological techniques. Literature [9], [10] have shown that it is possible to predict the personality of a user from his online behavior and social media activity. Following these findings we will make an effort to construct fake social media profiles that would advertise psychological vulnerabilities.

We believe that if there is no direct way for an attacker to contact a high profile employee he will target a person from his social or working environment which combines specific personality characteristics. Our second hypothesis is that if we create a fake person which carries the characteristics the attackers look for then they will choose to approach or communicate or attack to that fake person. That person in the working environment is a fake employee and is acting as a social media honeypot to help us detect spearphishing.

Next, we will present findings from research work related to phishing and spearphishing and what the phishers look for in a user's personality. We will also present a correlation between personality and online behavior and the language the users prefer to use in each case. Since the attackers harvest personal information in social networks like Facebook, this platform is chosen to advertise the characteristics of the fake personality.

3. Background

Nowadays it is commonly accepted that the weakest link in the cyber security chain is the computer user [1]. Honeypots are used to protect information systems but in our case we want to protect the users from being deceived before they reveal confidential information about the organization's information system or other employees. In this section of the report we will present a U.S. Patent of Uri Rivner and Idan Aharoni [17] which is a honeypot system against phishing and spearphishing emails, the recent CIA's Director Email hack [2], the "Robin Sage" [3] project presented in BlackHat USA and an interesting experiment which created a fake business and bought amazing online reputation [18].

3.1. Protecting Electronic Assets Using False Profiles In Social Media

The United States Patent US 8,856,928 B1 is a project for "Protecting Electronic Assets Using False Profiles in Social Media" by Uri Rivner and Idan Aharoni. In this project an administrator is creating fake social media profiles that correspond to fake employees of a company. The profiles contain details about where this person is working, the position, education etc. but there is no visible email. An intruder might fabricate an email in a typical corporation format of name.surname@corporation.com from the information harvested from the social media profile and try to contact the fake employee. Then, the receiver creates an alert that notifies the administrator. He then, checks the email, stores the FROM address, REPLY TO address to a database and every email coming from that address to the mail server is dropped before reaches employees mailboxes. That way the employees are protected from a phishing or a spearphishing attack.

The above concept has similar parts with our idea of creating fake social media profiles for fake employees in order to deceive the intruder. When an email arrives to the corporate mail server an evaluation process starts and will determine if the emails coming from that address will be dropped before they reach the mailboxes of the employees. There is also an important difference. Uri and Idan's honeypot is a system that is waiting to be discovered by the intruder. Our system is acting like calling the intruder to attack instead of waiting to be discovered. The call to the intruder is performed through social media networks by using psychological profiles, similar to the ones, we assume, that the attackers look for in order to launch an attack.

3.2. CIA's Director email hacked

Very recently was published an interview on Wired[5] of a teen hacker who, as he stated, with the help of two other persons they hacked in CIA's Director John Brennan AOL email account using social engineering techniques.

The hackers made a background check on John Brennan and found that his mobile number belong to Verizon. His email account was in AOL. They discovered that in order to reset AOL's account credentials they needed Brennan's personal info like account number, his four-digit PIN, the backup mobile number on the account, AOL email address and the last four digits on his bank card.

The hackers contacted Verizon pretending to be Verizon workers and they needed to provide support on a customer. They provided a fabricated Vcode (Verizon employee's number) to the Verizon employee and they received all Brennan's personal info's that were needed for the job. The interesting part which is not explained in the interview is how they managed to fabricate the private Vcode number which authenticates them as Verizon employees.

After the hackers got Brennan's personal info they contacted AOL. AOL asked some security questions like name, telephone number associated with the account and the last 4 digits of Brennan's bank card and they reset email's account password. The hackers managed to take from Brennan's email personal and official documents. Some of them were published in their Twitter account as proof of their success.

Brennan tried for 3 days to take control of his account but he couldn't because the hackers were resetting the password again until he was notified that he was hacked. After that he deleted his AOL email account.

3.3. Getting in Bed with Robin Sage

One experiment which is similar to our work is the "Getting in Bed with Robin Sage" by Thomas Ryan [3]. It was presented in BlackHat USA 2010 and the goal was to exploit the "fundamental levels of information leakage". The experiment gives a strong

5 "Teen Who Hacked CIA Director's Email Tells How He Did It" by Kim Zetter 19/10/2015, http://www.wired.com/2015/10/hacker-who-broke-into-cia-director-john-brennan-email-tells-how-he-did-it/

argument that the weakest link in information security is the human factor.

For a period of 28 days a fake profile was created in different social media. The person who seemed to own these profiles was an attractive young woman, 25 years old, named Robin Sage. Her pictures were taken from a pornographic website. Her job title was "Cyber Threat Analyst" at Naval Network Warfare Command, and claimed to have 10 years experience in the field. She graduated from Massachusetts Institute of Technology (MIT).

In these 28 days experiment she managed to connect with a few hundreds of persons from government entities such as NSA, DoD and Military intelligence groups. By connecting with well respected specialists her credibility rapidly rose. Then it was easier for other people to connect with her by accepting her friend request or by asking her to accept their friend request. The thinking was simple, "since she is a contact of respected person A then I could trust her too" but foolish.

During the experiment Robin was asked to telephone conversations, she was offered job positions and free tickets to conferences. One NASA Ames Research Center lecturer offered to share his paper with her before even published. But some managed to find a work around and reveal her fake identity.

No one doubts the education, the skills and how focused they are on their duty all these people who were fooled by her but it shows that nowadays is probably easier to hack a person's perception at a given time under certain conditions than some machines.

3.4. I created a fake business and bought it an amazing online reputation

A story was published on Fusion[6] about a person who created a fake company and found ways to make it very popular. The company was named "Freakin' Awesome Karaoke Express" or (F.A.K.E.) but no one paid attention to what the acronyms actually meant. That company claimed to have a great looking truck that would deliver the best music and equipment for organizing karaoke party. The author explains that she wanted to investigate how the fake reputation economy works. She found an online marketplace

6 "I created a fake business and bought it an amazing online reputation" by Kashmir Hill, 15/09/2015 http://fusion.net/story/191773/i-created-a-fake-business-and-fooled-thousands-of-people-into-thinking-it-was-real/

that would create reputation for her business. For $5 she could get “200 Facebook fans, or 6,000 Twitter followers, or @SMExpertsBiz to tweet about the truck to the accounts of 26,000 Twitter fans”. For $5 more she could get Facebook, Google, Amazon and Yelp[7] reviews. She created a web page, a Facebook and a Yelp profile and a Twitter account.

When the good reviews and comments online started to multiply the owner received emails and messages on her answering machine because people wanted to hire her. One more star review in Yelp increases job by 5-9% according to Harvard Business School’s Michael Luca.

[7] http://www.yelp.com

4. Related work

In this section we will present related work that helped us through our project. An effort has been made by researchers to develop an email honeypot system to automatically receive, analyze and archive spam emails [19], others developed new honeypot architecture to attract sophisticated attacks to secure systems by deluding even professional attackers with high knowledge about the target environment [16]. Two other approaches propose to transform a real banking system into a honey-net in order to automatically detect attempts of stealing money by phishers [20], [21]. An interesting project [22] studies the impact of publicly available personal information on social networks in phishing and spearphishing attacks. A different research tested the impact of anti-phishing tool performance on attack success rate [23] and two other efforts analyzed the textual content of the emails and attachments [24],[5] sent to possible victims. A report [7] providing a survey of literature relevant to the attribution of spear phishing attacks is made by the Australian Department of Defense. This report is based on profiling the attacker, it is dealing with the problem of attribution in spearphishing and is proposing to analyze the different phases of a spearphishing attack as 4 separate crime scenes which they produce their own evidences. In the proposed attribution approach it is suggested the use of proactive measures such as decoy elements to collect useful information during Target phase.

The first step of a spearphisher is Target. In this phase information are collected about the person in target. This information can be collected by a trusted insider (employee) or an entrusted outsider or a direct communication [7]. The direct communication is a part that we will be interested during our project because as it is mentioned in [7] a proactive approach would involve the organization planting ahead of any targeting activity a wide range of decoy elements. The direct communication can be a phone call, an email, and IP addresses visiting our website. In the absence of the above, harvesting publicly available personal information from social networks is the next solution. Target phase is far from the first incidents that might notify us of an attack, therefore it is possible that the phishers might make use of real IPs, phone numbers etc. That is why it is very important the logging of all information. An experiment in [10] showed that three in six identified groups were responsible for up to 86% of all phishing incidents, so, from this phase of attack we could gather information about the “class of adversaries an adversary belongs to and the type of victim as well as types of information the adversary is after”.

In the area of cyber-psychology and the user's behavior in Facebook the findings of two very interested papers [9] "Personality and Patterns of Facebook Usage" and "Personality, Gender and Age in the Language of Social Media: The Open-Vocabulary Approach" [10] will be used to construct the main element of this project which is a false Facebook profile for a fake employee. The researchers have shown that personality traits can be predicted through the online behavior of a user. They present the correlation between their personality trait and the number of likes, groups joined, photos uploaded, how many times have been tagged by others etc while in [10] they show the language style and word clouds they use according to personality, gender and age.

In [25] we find information about which parts of a Facebook profile the users tend to use in order to make judgments on others. According to the same paper, female targets have a higher probability to disclose information on relationships, proud or embarrassing things they did, good or bad times they had and things that make them feel great or crazy. On the other side male targets are more likely to disclose on Facebook political views, impressions about people, art, movies and websites.

In the area of cyber-psychology and social media like Twitter and YouTube [12], [13] the researchers presented ways to identify possible malicious insiders according to their online behavior. The researchers correlate the language used in the comments of YouTube videos to predict if a user has negative feelings against law enforcement and authority [12]. A similar research on Twitter accounts are presented in [13].

In the "Social Phishing" research paper [22], a demographic between the people that are susceptible to phishing according to gender or major they studied has estimated the percentage of people being susceptible to phishing to 77% Females and 65% Males. Similar findings are presented also in [5]. The attack is even more successful if the email comes from the opposite gender with greater success on males. In addition, in the same research a control group of University students received the same spam message from an unknown person with University email. Only the ones who majored in education had high success phishing rate of 50% but when the attack exploits information harvested from social networks the success rate is very high. Students majored in Business from 19% success rate went up to 72% with the help of the information found on social media. The highest gap is coming from Science and rises from 0% to 80% when the attack exploits personal information from social networks.

Technology majors such as computer science, informatics and cognitive science are the least vulnerable [22].

In [19], has been found that the best place to advertise emails for phishers are the online guest books. It was assumed that is even better than adult content websites. During our research in this project it was found that this is not valid anymore because online guest books have changed the functionality they had some years ago. There are no visible emails for an attacker to harvest. Even large hotel chains do not have online guest books but they scan and publish on their websites pictures from the physical guest books.

In [20], it is presented that victims usually reply immediately after reading the phishing emails and if someone replies after a day that is an indicator for the phishers that is not a good victim. In addition on the same matter in [7] is presented that the phishers make use of different techniques in order to extend the lifetime of a phishing website. The different types of phishing websites is Non-Rock, Rock domain, Rock IPs, Fast Flux domain and Fast Flux IPs as shown in the table below [7].

	Mean Time(hrs)	Median Time(hrs)
Non Rock	58.38	20
Rock Domain	94.26	55
Rock IP's	124.9	25
Fast Flux Domain	454.4	202
Fast Flux IP's	124.6	20

Table 1. Lifetime of different types of phishing websites [7]

Non-Rock phishing website is the simple phishing website that impersonates, for example, a legitimate banking website to deceive a possible victim while Rock phishing "is the phishing tool and the entity that publishes the toolkit[8]". The Rock phish toolkit gives the opportunity to publish under the same server different phishing websites like "Banks" and different folders[9]. The fast-flux technique is when a domain name like

[8] Rock phish definition, http://searchsecurity.techtarget.com/definition/Rock-Phish, 17-12-2015

[9] F-Secure have published a demonstration of a live Rock Phish Kit "Several "banks" are hosted on one server" on YouTube, https://www.youtube.com/watch?v=6NviimO64qA, 30-12-2015

www.example.com has multiple IP addresses assigned to it. These IP addresses are changing very often and it is possible that 2 pc's requesting the same phishing website with a small time difference to have the content delivered from different infected machine [26].

The assumption that the victim in order to be a good victim should respond right after reading the email might apply to Non-Rock phishing attack because more or less the lifetime of the phishing website is 1 day. So in this case the phishers are in a hurry. But in Fast Flux Domain phishing attack the phishing website has 202 hrs median lifetime and almost double amount of Mean lifetime. That might mean that the phishers are willing to wait for more victims to come as long as they feel secure. Of course, the phishers still might be interesting on the fast answering victims and by extending the phishing website lifetime their goal is to create a bigger pool of victims.

In 2015 Facebook had almost 1.5 billion active users [27]. Active are the users logged in to Facebook in the last 30 days. Chief Executive Officer Mark Zuckerberg announced last year that "the average U.S. consumer spending 40 minutes on the social network each day" [28]. People seem to be very interesting of what others think about their online appearance and they spend a large amount of time forming their Facebook profile expecting to receive a big number of "likes" on their uploaded photos or positive comments on their status updates. Dr. Michal Kosinski and his team in their paper "Personality and Patterns of Facebook usage" [9] show how to predict the personality of a Facebook user through the theory of the five traits of personality according to the number of status updates, number of published photos, events created and groups joined, "likes", number of friends and how many times the user have been tagged in photos by others. It is also proposed a correlation between personality traits and Facebook profile features. For example, a person high in neuroticism "likes" a lot of his friend's postings such as photos or status updates. In "Personality, Gender and Age in the Language of Social Media: The Open-Vocabulary Approach" [10] it is presented a correlation between personality traits and words used by Facebook users according to Personality trait, Gender and Age. They are presented in word clouds so someone can see which words are used more frequent and which one has stronger correlation.

Research paper [29] is dealing with the way a user is accepting someone's friend request and how a user is choosing a person to add it as a contact in social media. The

researchers conducted an experiment by creating fake Facebook profiles to study the behavior of the users and it was found that women attract more friends than men and celebrity profiles attract more than non celebrity ones. Celebrity profiles are less successful in adding friends because apparently the normal user assumes that there is something phishy in that request. On the other side celebrity's profiles receive a large number of friend requests by staying idle and they can easily be used as honeypots to collect user data.

The information gathered from the aforementioned literature will be used to construct our fake Facebook profiles for the needs of this project. Carefully choosing gender, age, personality trait, pictures and by creating a realistic and promising background for our fake profiles we believe that we can create a social media honeypot to detect attackers like phishers and spearphishers.

5. Related work II - Social Engineering

In a new type of eCrime called Advanced Persistent Threats (APTs) the goal of the attacker is to compromise target systems for long periods of time. No matter how sophisticated are the attacks the first stages very often are done using social engineering techniques like phishing and spearphishing [5]. Social Engineering is the process of deceiving people in order to disclose classified information.

5.1. Phishing

Phishing is a form of deception in which an attacker attempts to fraudulently acquire sensitive information from a victim by impersonating a trustworthy entity, like personal information, intellectual property, financial data etc [22]. Spear phishing is the targeted phishing, is the attempt of an attacker who has studied the possible victim and is using information about him or his social environment in order to gain his trust and finally distract large amounts of money or valuable personal information. It is performed in 4 steps [7], Target, Lure, Hook and Catch. The type of spear phishing which exclusively targets groups of high level executives in an organization is called Whaling. The reason Phishing and Spearphishing is such a big problem is its huge profit margins, it is easy to perform the attack and difficult for the law enforcement to find the phishers [22].

Thus, the goal of the social engineer is to affect the target's emotional state and the ways to do that are only limited by the imagination and creativity of the attacker. The attacker could visit the area he wants to collect information, the offices of a company and search for names, positions, telephone numbers, and email of a target. At the same time the attacker could search the Internet and social media to find plenty of personal information for the target and its colleagues which they will be used in a psychological game against the target. That type of psychological games focus on persuasion, ingratiation, conformity and friendliness [4]. Research has concluded that "pressures of conformity, compliance and obedience cause people to change their behavior" [4]. That means that someone could pretend of being a high profile employee and without providing any identification to make the other side to reveal valuable information.

Phishing is performed in two different approaches, a computer based and a human interaction based [4]. In the computer based approach the attackers will send an email to the target and will try to convince him to click on a link that will redirect them at a

website looking like a Bank institute or Email provider asking for credentials or other personal information. When the victim submits the information to the phishing website the data are sent to the phisher. Another way of collecting the desired information is by infecting the computer of the target with malicious software that records the keystrokes and sends them back to the phisher. The human interaction approach is based on deception and impersonation. A phisher pretending to be a person with authority, a colleague, a manager or a trusted third party is contacting the person in target and by establishing a trusting relationship collects the desired information. In all cases urgency is the key feature that the social engineers are using. They do not want the target to start thinking, analyzing or even cross checking the validity of the story told by the social engineer.

5.2. Spearphishing

One other way to social engineer a target is Spearphishing. Spearphishing is the targeted attack on a future victim making use of all kind of information that are able to be collected from the place of work, friends or social networks in order to deceive the target and extort financial, intellectual or other gain. Spearphishing is conducted in 4 phases, Target, Lure, Hook and Catch. In the Target phase there is not strong evidence that someone is looking information to conduct an attack on a target. That is why it is the most difficult phase to identify.

The Target phase is the procedure when an attacker is searching for a victim to exploit. The attacker might be a single person, an organization or a State. The intruder knows very well what information they want to collect from Private or Public Sector. Then they start researching which person has that information and how to contact him. After the person is targeted then a social engineering operation starts to collect any relevant information on that person. This information can be gathered by an insider, by an entrusted outsider or by a direct communication between the phisher and the target [7] and if none of the previews are successful, social media are always a place to harvest much personal information.

Next phase is Lure. In this phase most of the times an email is sent with a malicious attachment or a link leading to a malicious website. The email seems to be from a person he trusts or an organization he is working with and receives usually emails from. The phishing email is addressing to the recipients in an abstract way like “Dear

customer ...” but in spearphishing the email is containing the name or/and the surname of the recipient to make it more believable, like “Dear John ...” The story in the email is convincing enough to lure the target to open the attachment or to click on the link. The target that is doing what he is asked to do in the email is probably hooked and moves to the next phase [7].

Hook phase starts when the malicious software is installed in the target’s computer. The software records what the target is typing or might make changes to the computer system helping that way the adversary to gain access to the system. Another way is when the target is clicking on a link sent by email and redirects him to a website which imitates the legitimate one in look and feel. The attackers are using the original logos and banners to successfully deceive the targets. Some other way is to use realistic looking URL’s like www.paypai.com or www.paypa1.com with I or 1 at the end instead of L [7].

When the malicious software or the fake website has collected the personal data of a target then they have to be sent to the adversary’s site in a secure way to minimize the chances of being caught. This is the Catch phase and usually is done through a covert channel. The phisher could send the data through public repositories such as newsgroups, chat rooms etc and go there later and collect them without being exposed.

Research in [7] suggests that the offender’s profile can be focused on 3 elements. Motive, personality and behavior. Motive is what drives the offender to proceed to this attack. Profiling an attacker is related to psychological research and appears to have two approaches [7]. The one is based on personality traits of the attacker and the second one says that the personality is influenced by the situation. That means that maybe a poor kid with great skills when in need or when it sees an opportunity to have quick and easy large financial gain, will decide to conduct an eCrime. On the 13th of October, Mikko Hypponen of F-Secure was a guest speaker at IT College in Tallinn and he referred on the same problem in a similar way. He said[10] “people with skills without the opportunities will have to make a living. If they can’t make a legal living they will make an illegal living.”

[10] Mikko Hypponnen at Estonian IT College, http://www.itcollege.ee/blog/2015/10/13/avalik_loeng/ , 13-10-2015.

Spearphishing is a difficult problem to tackle because it is using deception which means that even people educated in the context of information security might be victims. The incident with CIA's Director Email hack shows that training about how to be protected from this type of attacks is not always enough. One way to defend or help the investigation of a spearphishing attack is taking proactive measures by planting ahead of the attack decoy elements in a company or organization. One example is Bowen's decoy system [7]. Bowen's decoy system suggests having a network-based decoy system to protect against eavesdropping and a host-based decoy system to protect against data exfiltration carried out by malware or insiders. The proposed honeypot in our project is a type of decoy system that aims to protect against data disclosure. The honeypot could interact with the possible spearphishers providing information to the company about what the attackers are interested in and possible reasons for the attack. The goal is to create an attractive model of an employee that the attackers will have the opportunity to approach and study him through his online behavior. The model will be constructed according to what research suggests for the personality of a victim.

6. Related work III – The victim's profile, attack vectors, persuasive techniques, triggers and cyber-psychology

In this section we will present a demographic of the people susceptible to phishing and spearphishing according to age, gender, education and anti-phishing training methods. We will also present what are the attack vector, the persuasive techniques and the triggers used by the attackers to accomplish their task.

Research paper [22] has estimated the percentage of people being susceptible to phishing to 77% for Females and 65% for Males. Similar findings are presented also in [1]. The attack is even more successful if the email comes from the opposite gender with greater success on Males. In addition we see in the same research that in a control group which university students received the same spam message from an unknown person with university email, only the ones who majored in education had high success phishing rate of 50%. When the attack exploits information harvested from social networks the success rate is a lot bigger. 19% of students majored in Business were phished but with the use of social networks the success rate went up to 72%. The highest gap is coming from Science major which goes from 0% to 80% when the attack exploits personal information from social networks. Technology majors such as computer science, informatics and cognitive science are the least vulnerable [22].

Another interesting finding is that age plays a significant role and people aged 24 years and less are more susceptible than older ones because young users are known to have high level of Agreeableness [1]. In the same research paper is presented a study that shows that females are more agreeable than males which also correlates with why females are more susceptible to phishing attacks. Female targets have a higher probability to disclose information on relationships, proud or embarrassing things they did, good or bad times they had and things that make them feel great or crazy. The words that have stronger correlation and frequently used by females are, "excited", "shopping", "love you", "my hair", "happy", "wishing", "family", "wonderful", "cute", "mommy", "her". On the other side male targets are more likely to disclose on Facebook political views, impressions about people, art, movies and websites. The words that are more correlated and used by males are types of bad language, "f***", "wishes",

"himself", "ps3", "xbox", "government", "battle", "football", "my girlfriend", "my wife" and "black_ops".

	Disclosed information	**Words used**	
Males	Political views, impressions about people, art, movies and websites	f***, wishes, himself, ps3, xbox, government, battle, football, my girlfriend, my wife, black_ops	Malicious insiders are Males by 75%.
Females	Relationships, proud or embarrassing things they did, good or bad times they had and things that make them feel great or crazy	excited, shopping, love you, my hair, happy, wishing, family, wonderful, cute, mommy, her	77% of Females are susceptible to phishing. Females are more agreeable than Males.

Table 2. Information disclosure according to gender

When it comes to Age factor, younger people tend to speak with bad language and use words about college, parties, exams, assignments, etc which are representative of their life. After the age of 23 years old, research suggests [10] that people talk more about work, companies, weddings and money. After 29 years old the use of "I" decreases and the use of "We" is becoming more and more frequent. Higher ages use fine language in their topics. The following table summarizes these results.

Normally the training methods provided by companies are of non-embedded type. An anti-phishing non-embedded delivery method is when the employees are first trained and then tested on the knowledge they attained. An embedded delivery method is more complex, it requires a controlled phishing attack against the people to be trained and then the education part starts. That way it is shown in [1] that the employees retain more knowledge.

From the above, the profile that we assume will be more attractive to a spearphisher would be of an 18 to 24 years old female with no anti-phishing training or at most a non-embedded type of training. The education that should have is preferably major in

Business, Science or Education. On the other hand, a person that is less attractive to spearphishers is a male, older than the age of 25, with training in anti-phishing performed with embedded delivery method and with education background Computer Science.

	Highly Susceptible	Less Susceptible
Age	18-24 years old or less	25 years old or more
Gender	Female or a Male contacted by a Female	Male
Anti-phishing Training	No training or Non-Embedded	Embedded
Education	Education, Business, Science	Computer Science

Table 3. Characteristics of high and low phishing

6.1. Attack vectors

An intruder like a phisher or a spearphisher is trying to develop a trusting relationship with the target as we have seen so far from the various studies. The psychological attacks use the following attack vectors focusing on persuasion, impersonation, ingratiation, conformity and friendliness. Pressures of conformity, compliance and obedience cause people to change their behaviors and the social engineers have learned to predict responses to these pressures [4].

6.2. Persuasive techniques

It is found that the social engineers use eight persuasive techniques [4], authority, urgency, tradition, fear/threat, attraction/excitement, pity, politeness, and formality. The definitions of the persuasions are following as they have been presented in [4]:

1) Authority: When the attacker is using convincing statements in order to create legitimacy, trust, and credibility. They claim to be some high profile employee of a

bank or university or other institution that will put pressure on the target to comply with the instructions.

2) Pity: It is used to create emotions of sympathy and charity in the messages.

3) Tradition: An appeal to ideal values such as honor and legacy commonly recognized by the public.

4) Attraction: It is used in order to attract the target. It is used for the Lure phase of a phishing attack. It could refer to job opportunities or big prices etc.

5) Urgency: Urgency in a message is used to force the target to comply quickly to the instructions of the attacker without giving the time to correctly evaluate the situation. Usually it is a threatening message like “respond to instruction otherwise your bank account will be disabled”.

6) Fear/threat: Fear is used in the same way as urgency, to impact the emotional state of the target and make him decide without giving proper thought of the situation.

7) Politeness: It is used for deception, to distract the mind of the target from the real intentions of the attacker.

8) Formality: Formal way of speaking or writing is used to create the impression that the person behind the letter or the phone call is who he claims to be.

The frequency the persuasive techniques are used according to the findings in [12] are Authority 100%, Politeness 74%, Urgency 71%, Formality 55% and fear/threat 41%.

6.3. Triggers

A phishing email can be categorized according to the trigger used in email body. A trigger can be defined as the main reason why a target will respond to the email. For example as trigger can be used an account update, account verification, account suspension etc. In [4] it is found that the phishers use eight triggers. Alert/warning/attention, account verification, Invalid login attempts, security update/grade of an account, account suspension, purchase confirmation, general update/grade of an account and identity verification. The 4 most used triggers in a

percentage between 15 and 20% are Alert/warning/attention, account verification, invalid login attempts and security update/grade.

According to the above findings, if we would like to create a convincing story and contact a target in order to be the victim of a phishing attack we would use an email provider personnel (Authority) to ask in a polite way (Politeness) a target to comply quickly (Urgency) to the change of his email password (Trigger: Security update) at a proposed link (Hook).

6.4. Cyber-psychology

A proof that the spearphishers have knowledge of cyber-psychology is that they make use of different psychological techniques [4] in order to impact the emotional state of a victim and influence them to reveal personal information. For that reason a big part of this project focuses on cyber psychology. We will try to show among others, how the personality is defined according to the theory of the five traits of personality, how to take advantage of this knowledge and turn it against the spearphishers.

Personality-psychological Internet research is dealing with the relation between the personality of an Internet user and how that person is using the Internet, how impacts their online communication and interaction. That means that according to the findings of the personality analysis of a person we could have an idea if that person is a possible online victim or not. Personality-psychological Internet research is using mainly the theory of trait psychology. The trait-psychological Internet research is trying to relate the five traits of personality and the online behavior [4] [9].

The five traits of personality are **Neuroticism, Extraversion, and Openness for experience, Agreeableness and Conscientiousness**. The definitions of the trait psychology's components in relation with on-line and Facebook activity are given below according to [9]. The research conducted in [9] and [10] reflects the society in United States. We will use the findings in a European audience due to the fact that people in the western mentality countries share common trends and ways of thinking. Despite the language differences we will assume that researchers would have had similar findings if the aforementioned research was conducted in Europe. The same assumption will apply on another research [12], [13] on how to predict malicious insiders from their online

behavior which was conducted by studying Greek YouTube and Twitter users.

"**Neuroticism**, is often referred to as emotional instability, is a tendency to experience mood swings and negative emotions such as guilt, anger, anxiety, and depression. Highly Neurotic people are more likely to experience stress and nervousness, while those with lower Neuroticism tend to be calmer and self-confident.**"** If we translate that in Facebook behavior it means that a user with high scores in Neuroticism is making large number of "likes" [9] and that is happening because Neurotic people are experiencing an isolation from offline social activities and therefore they hope that by liking many things in Facebook, their online friends will reciprocate providing the psychological support they hope for. Neurotic people have a medium number of Facebook friends of around 200. People with much lower number of online friends or much larger than 200 they usually have small scores in Neuroticism. Neurotic people also tend not to participate in many Facebook groups. The number of people someone is interacting with in online groups becomes very big and the people in the groups are basically unknown therefore a person high in neuroticism prefers not to join in. Neurotic people prefer to be liked by people they know, and that is why the number of online friends is not exceeding the 200. In [10] is mentioned that Neurotic people use more negative emotion words. They tend to use more first-person singulars like "I, me, mine" [30] and more acronyms. When anger is noticed in language is probable that this person is high in neuroticism. The language they prefer to use contain words like "f***", "depression", "hate", "alone", "stressed", "sick of", "worse", "annoyed", "pissed", "dead". On the contrary people low in Neuroticism use words as "great", sports related and "work out", "beach", "praised", "blessed", "God", "weekend". Finally, **neuroticism is positively correlated with people replying to phishing emails** [5].

"Extraversion measures a person's tendency to seek stimulation in the external world, company of others, and express positive emotions. Extraverts tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative; do not mind being the centre of attention, and **make new friends more easily**. Introverts are more likely to be solitary or reserved and seek environments characterized by lower levels of external stimulation.**"** In Facebook terms extraverts like to express their sympathy on others by liking their pictures, comments or status updates. That is why they participate in many online groups exchanging views with people they don't know. Extraverts have many

friends and they can befriend a lot easier than others, thus they can be contacted more freely by an offender. Research suggests [10] that extraverts use longer words; they also use frequently words as "party", "girls", "can't wait", "chillin", "love you", "night with", "baby", "weekend" while introverts use "anime", "manga", "pokemon", "computer", "windows", "program", "episode", "internet".

"Openness to experience measures a person's imagination, curiosity, **seeking of new experiences** and interest in culture, ideas, and aesthetics. It is related to emotional sensitivity, tolerance and political liberalism.**"** People having high scores in Openness they have large numbers of "likes", participating in big number of online groups and they post a big number of status updates too. According to [9] they seek new experiences, they like adventure and new or unusual ideas. These characteristics might be of interest of an attacker. A spearphisher who is looking for a possible insider to exchange information for money or adventure might want to contact a person who has the tolerance to talk about unusual ideas. People low in Openness talk more about social processes, use words like "Mate, talk, they, child" [30] and more quotations [10] while the ones having high scores in Openness use words like "music", "art", "writing", "dream", "universe" and "soul".

"Agreeableness measures the extent to which a person is focused on maintaining positive social relations. High Agreeable people tend to be friendly and compassionate, rather than cold or suspicious. They are **more likely to behave in a cooperative way**, **trust other people**, and adapt to their needs.**"** That means that they seek for other people's "likes". So, when they express themselves they do not do it as freely as others but they take under consideration the feelings and likes of the people they associate with. According to [9] agreeable people publish photos with others and they receive many tags by others. Thus, in Facebook terms, users scoring high in Agreeableness they have high numbers of tags by others and they have small numbers of likes on others. It seems that trait-psychological Internet research has the biggest difficulties in predicting that type of personality. According to [9] people having high scores in Agreeableness are more likely to compromise and may be more gullible. Their trust on others good intentions make these people desirable by Phishers and Spearphishers. People with high Agreeableness seem to be perfect to play the role of a "blondie", the person who will make the "click" and give access to an outsider. People high in Agreeableness use more

articles like "A, an, the" [30], and first-person singulars like "I, me, mine" [10]. They also use words such as "excited", "amazing", "wonderful", "prayers", "blessed", "lord", "awesome", and “Merry Christmas”, “thanksgiving”, “Christ” and “thank you”. People low in Agreeableness use bad language, words that correspond to hostile attitude and substance abuse [11].

“Conscientiousness measures preference for an organized versus spontaneous approach in life. People high on conscientiousness are more likely to be well organized, reliable, and consistent. They enjoy planning, seek achievements, and pursue long-term goals. **Low conscientiousness** individuals are generally more easy-going, spontaneous, and creative. They tend to be more tolerant and **less bound by rules and plans**.**”** High scores in conscientiousness are translated in small number of likes and small participation in online groups. According to [9] that is happening possibly because these people believe that spending time on Facebook is waste of time and distraction from other profitable activities. The interesting finding is that people with high conscientiousness are posting a large number of photos. The tendency they have to upload large number of photos though, might give the opportunity to an attacker who managed to be close to them to study from their photos their everyday routine, what they like and which are the people they are connected to. In [10] is mentioned that people low in conscientiousness use "more words signaling distinctions". People high in conscientiousness prefer to use words such as "work", "ready for", "blessed", "wonderful", "long day", "success", "relaxing", "thankful", "vacation", "workout", "great day". The ones in the opposite direction use words that reflect being disorganized, irresponsible, impractical, careless such as "f***", "pokemon", "youtube", "facebook", "bored", "anime", "gay" [11].

In [12] and [13] researchers claim that they can predict people possible to act as insiders according to their online behavior. In the first paper they analyze the comments on YouTube and on the second the lines written in Twitter. The research conducted in the micro-world of Greece and we will use the findings with two assumptions. Given the fact that the majority of videos posted in Facebook are YouTube links then if we can predict an insider’s tendency from the comments he makes or others commenting under the content he uploaded on YouTube then that should apply to Facebook as well. Thus, we will consider as possible insider a person who makes comments in Facebook against

law enforcement and authorities and posts YouTube content which contains that kind of comments. We will also use the findings from [13] which they come from analyzing Twitter and we will assume that they apply to Facebook as well because we believe that is important to concentrate on the tendency behind a person having a specific online behavior and not the media he is publishing his ideas.

It is observed that insiders develop the “revenge syndrome” and anger against authority figures. A demographic shows that 75% of those with negative feelings against law enforcement and authorities are males [12]. In [13] is mentioned that the insiders have high scores in the personality trait of narcissism. Studies have shown [31], [32] that people who are narcissists use Facebook to self-promote in ways others (who are not trained) can identify. Some of the things that narcissists need as persons are leadership, authority and great exhibitionism. That is why they choose “glamorous, self-promoting pictures for their main profile photos”. Narcissists tend to have a very large number of online friends, update about their achievements, appearance and they do not like to update about intellectual topics. In general they prefer quantity instead of quality.

7. False social network profile

In order to proceed with the creation of false Facebook and LinkedIn accounts we need to answer to some simple but important matters that will determine the nature of our honeypot. Can we have a fully automatic system or not? Do we need to have more than one fake profiles to create a realistic looking fake profile? What value the honeypot will give to the organization and why bother developing and maintaining such a project. Finally, we will propose how to create a fake profile to meet the requirements of an organization and offer an additional phishing and spearphishing protection.

7.1. Online life

Online personality cannot exist autonomously, it needs to be created by online life and it has to look in this project as real as possible. The Oxford dictionary defines life as "The condition that distinguishes animals and plants from inorganic matter, including the capacity for growth, reproduction, functional activity, and continual change preceding death"[11]. What is online life and who can have one? Online life can have online death, has online activity, online life can be reproduced and has the capacity for growth. So, someone could say that online life is the life of a living being online. But, we can program a computer to create different social network profiles, have online activity (automatic status updating, commenting on others' post, posting pictures and uploading YouTube videos), growing by accepting and sending friend requests and facing online death by terminating the existence of those profiles. All the previous activities can be carried out by the computer itself. The status updates and the comments can be created by automatic sentence generator. To automatically post pictures or the created status updates and make comments one could make use for example Python or Google scripts. Because we must advertise a specific type of personality, the sentences need to contain words from a specific pool. That is obvious that will lead to problems with the created sentences that will not make sense or they will be irrelevant with the subject. One way to address this problem is to make use of single word sentences on status updates or

[11] Online Oxford dictionary, http://www.oxforddictionaries.com/definition/english/life

comments with words coming from a specific pool. In that case Facebook has a rule that when one is posting identical status updates after a few attempts they are banned. At the same time, when someone who is studying the character of a person who is posting only single word sentences, would understand that there is something strange with the particular profile.

To return to the original question of who can have an online life we conclude from the above that it has to be a real person. Thus, the honeypot that is proposed here cannot be 100% automatic, it should be at least hybrid with a real person behind it which he could manage more than one accounts and could make use of 3rd party applications to keep up-to-date all the profiles , but is important to have the supervision of what and how is written online. What does a real person need, who has Facebook profile, to have online life? The person needs to have online friends and communicate with them. Need to participate on different activities or events and expose himself online by publishing photos and/or videos. Then he will receive comments and "likes" about these activities and he will reciprocate to his friends on their own activities. One important thing that someone needs to have in social media in order to look real is history. The account cannot be blank or just recently created. When someone looks at an account that has no history most probably will think that there is something phishy with it. Facebook recently allowed companies and personal profiles to add content as backdated which changes a lot in the process of deception. Using the backdated posts we can create a scenario and presented in Facebook, a scenario that will involve the main character and the supporting crew.

7.2. Support crew

It is recommended to create a few more Facebook accounts to play the role of friends and family of the main character as supporting crew. That will help to generate comments, "likes" and events that will make the whole scenario to be more realistic. The support characters are very important because in the beginning they will be the first friends that will make the profile look decent. As we saw in the case of "Robin Sage" every new friend was accepting her based on the judgments of the previous people connected to her. So, the friends need to look real and decent with normal lives.

Another equally important element is the photos. The main character and some or all of

the supporting characters should have at least two photos of a real person claiming to be the owner of the profile. One approach would be to search the Internet for available photos to use in our Facebook fake profiles. In "Robin Sage" case, the creator of the fake identity stole a picture of a porn model. That might be ok for a 30-days experiment but in the long term someone might find where this image belongs to and expose the whole operation. At the same time there are legal issues coming from using someone else's photo without the owner's consent.

There are other reasons also why we should not use pictures from the internet. Someone might use advance facial recognition software[12] or the small world phenomenon[13] [33] might betray the cover. When the honeypot start to expand its friend base some of the connected people might know the person in the picture and expose the cover. If we search the internet for ways to detect a fake Facebook profile we can find many articles[14] proposing different tips. One of the tips is that we can navigate to Google images with chrome or firefox browser and drag-and-drop a picture from our desktop in the search area. Then Google will search for an exact match or pictures similar to the original. Same results we can have with an application called "FB Checker" which checks if the photos in a Facebook profile belong to the owner of the profile or they were taken from somewhere else. Thus, it is needed the organization to support the social media honeypot project with a team that would overcome the above detection ways and would be able to produce realistic looking photos for online use.

7.3. The value of the honeypot

[12] Links for advanced facial recognition software, http://www.imagus.com.au/, https://facedetection.com/software/, http://animetrics.com/forensicagps/

[13] The Small-World Phenomenon [33] proves that "two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances". Milgram in his small-world experiment used a source person in Nebraska with instructions to deliver a letter to a target person in Massachusetts. The times the letter was delivered successfully, the intermediate number of persons needed to forward the letter to the next hop was found to be between five and six.

[14] How to Reveal a Fake Facebook Account, http://www.wikihow.com/Reveal-a-Fake-Facebook-Account, "**Beware interconnected faking**. At one time it was probably reasonable to think that if someone had a group of friends interacting with them and vouching for each other, that that person must be real. Not anymore! There are increasing cases of one person running numerous fake Facebook accounts, pretending to be an array of different people, all vouching for one another and all trying to be friends with someone real! An excellent example is the case of Natalia Burgess (http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11161333), who wove a web of deceit and caused many young males to fall for her various aliases — all because she felt inadequately loved.[1] Sadly, impostors of this sort go to incredible lengths to create an array of fake accounts including other social media accounts and websites to give the impression that their fake personas are "real"."

The victim has a value for the organization and for the attacker. For the organization because the honeypot will be the alarm that someone or something is looking information on the field we want to protect. We can record times, methods used by the attacker or geographical areas that are interested on us. It is also valuable because if we manage to detect signs of phishing or spearphishing we can protect the personnel and in extend the organization faster than without the honeypot. If we share this information others will be aware too. At the same time the victim is valuable for the spearphisher because it is the tool to unlock the organization and collect the information needed. So, both sides need to provide support to the victim. From the organization's side the victim needs to be supported in order to look real at all times and the adversary's side needs to keep the victim in a friendly level. As we saw earlier, in 74% of the times the attackers use politeness in their communication. This proves that they want to be liked from the possible victim and this behavior shows signs of agreeable behavior from the attacker's side.

As it is mentioned earlier, the victim should look like a real person, to have the personality and the online behavior that will attract an attacker to take advantage of. An example of that type is the spam emails with some very obvious typing mistakes. Some people realize immediately that this is a fraud and that is why they are not good samples for victims while some others don't mind of the mistakes and reply to the email. That is how the possible victims are separated between the ones who are not good to be victims and the good to be victims. Normally, after the first email a new contact attempt will take place with the potential victims by email or other means. So, the personality is one of the key factors that will make the attacker approach or attack on the specific person.

7.4. Tools used during the project

During the project we used three online tools, Hootsuite, NetworkedBlogs and ApplyMagicSauce. Hootsuite[15] is a web based platform that someone can manage his social network profiles from one interface, can schedule messages, extract profile analytics, and automatically post RSS feeds from other websites. In its free version someone can manage up to 3 different network profiles in Facebook, LinkedIn, Twitter, etc. but it seems to have some problems with extracting profile analytics.

15 http://www.hootsuite.com

NetworkedBlogs[16] is a web based platform that someone can follow his favorite blogs in one interface and share on Facebook what he is reading. These two tools were used in order to save time managing Facebook accounts and to have a wide range of information offered to us every day in order to post on our fake Facebook profiles and create the impression of a real profile.

The third tool, ApplyMagicSauce[17] is a web based tool of the Psychometrics Centre of Cambridge University to predict the personality of a user according to his online behavior in Facebook evaluating the status updates, likes, tags, words used etc. This tool was used to calibrate our fake Facebook profile personality and get an idea of what a person with knowledge in psychology would infer about the personality of our fake profile.

7.5. How to create a fake social media profile

The initial and simple answer to that question is that we cannot in a legitimate way. The first rule in the terms of use of Facebook is "You will not provide any false personal information on Facebook, or create an account for anyone other than yourself without permission". The last two words make a big difference. They mean that there is probably a way to ask permission from Facebook to create false accounts. When we created the fake profiles, Facebook proposed to us to connect to an Estonian person with 600 friends. That profile belonged to a female person with no more than 3 photos. At the same time, when we created a LinkedIn profile the social network proposed to connect with 3 others. The Facebook proposal looked very realistic with photos and a name and surname while the LinkedIn proposals did not have pictures and their names did not appear correct. They were more like two words of random letters that did not make sense. We questioned ourselves "Why Facebook proposed that person with the 600 friends and not some of the other profiles which they have a few thousand of friends". We checked the information of the proposed person and we found no common interests, activities, schools studied, cities lived in the past, etc, that somehow we could be linked with. Thus, we thought that Facebook probably uses a type of social media honeypot to track false accounts created with an automatic way or accounts that their

16 http://www.networkedblogs.com
17 Apply Magic Sauce translates individuals' digital footprints into detailed psychological profiles, www.applymagicsauce.com

purpose of creation was phishing. It seems that fake accounts of that type were used in the case of F.A.K.E. business which bought a large number of Facebook friends and "likes" with $5. LinkedIn initial proposal seemed false with the first glance which makes us think that LinkedIn is targeting more the scripts but Facebook is targeting the scripts and the people who create an account manually and they are trying to create a huge friend base in a very small amount of time. So, we have a strong impression that social networks make use of social media honeypots to fight false accounts.

For that reason we contacted Facebook and LinkedIn by email and asked the networks their permission to create false profiles for security reasons. Facebook replied with an auto-generated email and they suggested to us to submit our question to the Facebook community. We did as they said but we did not receive any respond. LinkedIn replied in a few hours and they were very clear that "I completely understand your point, but creating fake accounts on LinkedIn will lead to the termination of those accounts. To avoid any further inconvenience, I recommend you not to create any fake profiles. If you have further questions, please feel free to reply to this message". Since we were given the opportunity to reply, we asked them, if a State or a Corporation asked officially the help of LinkedIn to create a social media honeypot, their answer would be different? They replied they will have to consult a different group. What we understand from these responds is that Facebook does not want to talk about it or our request did not find the correct person yet. On the other side, LinkedIn is negative in the beginning but maybe a different approach on a more official level might change their mind. We understand that the networks wish to convince the people that the information published in their media is reliable and credible. And that is why the phishers and spearphishers or any other person or agency wishing to harvest information about someone would pick these networks.

Since the networks seem to have a negative approach to provide officially help to a person to create social media honeypots how to continue the project? If someone proceeds to create a false social media account it does not seem to have legal implications if he does not do something bad. A person can do it like we did and the only implication might be the termination of the account if and when it is found. How a company could overcome the problem? A company A could outsource the project to a different company B that will deploy the social media honeypot and sell the services to

company A. If company B is registered in a country with low interest in cyber legislation, it is more convenient.

In our case the honeypot will be a fake employee in a fake small company. The employee will be connected to one high profile employee that the company would like to protect and detect spearphishing attacks against him. In Table 2 we presented the information that could be disclosed according to gender and on Table 3 it was presented high and low susceptibility to phishing according to gender, age and training. Combining the information from these two tables we decided that our fake employee will be a young woman and she will play the role of a personal assistant (PA). The high profile employee has the characteristic that is silent and he cannot be reached by anyone outside his circle of trust without contacting first his PA. Thus, the honeypot proposal to detect spearphishing is not just a few fake social media profiles that will be out there for the phishers to study them but also a proposal of how the high profile employees should connect and communicate with people they do not personally know.

The fake employee will be based on false Facebook and LinkedIn accounts providing information about her life and work experience. We have decided to use Facebook as the face of the honeypot because there have been studies that analyze the human personality based on what a person writes or likes on a Facebook profile [9], also, because the intruders use social media to collect information on their future victims. At that point we will test our hypotheses. We will manage the PA's profile according to the information gathered in Table 5 and we will try to give the impression of a 24 years old woman. We will create status updates according to the language style of a young woman and we will choose words and behave online in a way that corresponds to agreeable behavior. If we manage to simulate an agreeable online behavior and we receive a spearphishing email it will give a strong argument that our first hypothesis is correct. Spearphishers will not only possess psychology knowledge to craft convincing emails as we already know but they also perform the reverse process. They analyze the psychological profile of a victim prior to an attack through the information collected by social media in order to decide which psychological technique will be more successful. At the same time the second hypothesis will be true as well, because the spearphishers would have decided that a person with personality vulnerabilities placed next to a high profile employee is the best choice to contact and be used as stepping stone. So, the

steps we followed were to choose gender, personality trait, and use photos that will comply with the lifestyle of the scenario and the personality trait. Then we behaved online according to the information gathered in Table 5 about the specific personality trait.

Very often, after status updates or "likes" we were testing the impression the profile was giving by using the online personality prediction tool[18] of the Psychometric Center of Cambridge University. A summary of our fake employee's psychological profile is shown in Image1.The image shows that our fake Facebook profile is giving the impression that there is a 26 year's old female behind it. The bar showing the gender has on the left the indication Feminine and on the right the indication Masculine. The value 23% means that is 23% left to 100% of being feminine. Thus, this online tool predicts that behind the fake profile is a person with a probability of 77% being female. The digital footprint that we provided was the Facebook profile of the fake personal assistant. The predictions for the five traits of personality are explained as:

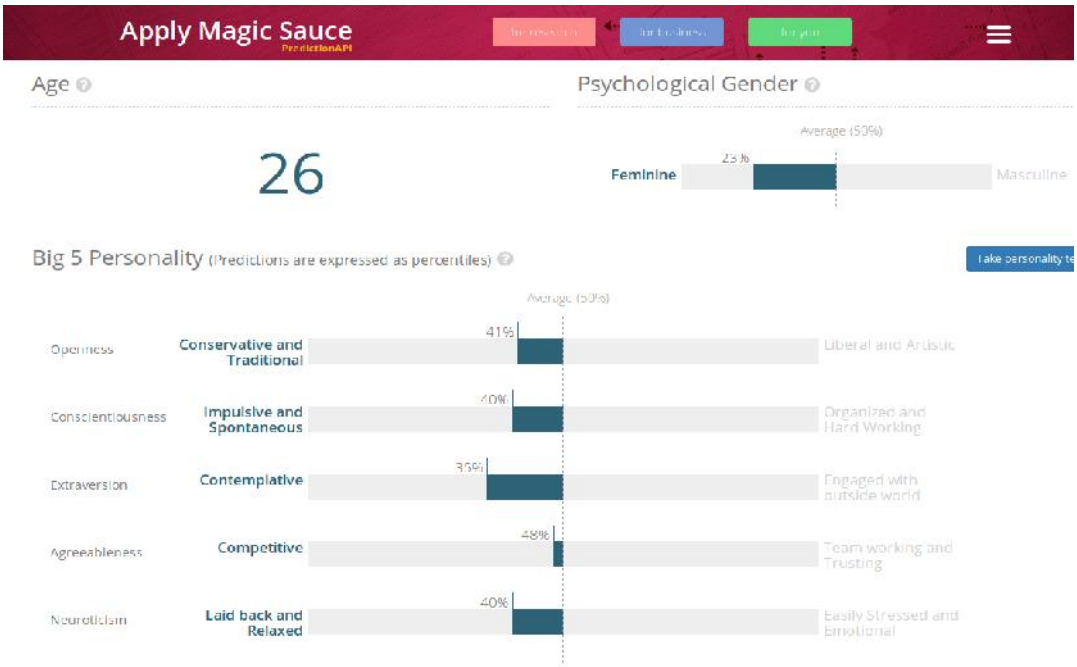

Figure 1. Personal Assistant's personality prediction

Openness 41%: It means that the profile shows that the owner is more conservative and

18 Apply Magic Sauce translates individuals' digital footprints into detailed psychological profiles, www.applymagicsauce.com

traditional than 59% of the population.

Conscientiousness 40%: The owner of the profile is more impulsive and spontaneous than 60% of the population.

Extroversion 35%: The owner is more contemplative than 65% of the population. This indicates that without the supporting “crew” it is very difficult to succeed extroversion.

Agreeableness 48%: The owner is more competitive than the 52% of the population. In this trait we were very close to be identified as Agreeable. To succeed that we needed more tags on other people photos. The role of the supporting crew is very important here as well.

Neuroticism 40%: The owner is more laid back and relaxed than 60% of the population.

From the above results it is obvious that we succeeded to represent the correct gender and age. We were almost agreeable but we failed in the extroversion part. We also failed to receive spearphishing emails that would prove our hypotheses. The reason we could not be as extroverted as we wanted is the lack of supporting fake characters which will play the role of the close friends of our fake employee. The reason we did not receive any spearphishing emails we believe is because the fake PA was placed next to a company director that no one has to gain anything from him. The fact that we used a fake company instead of a real one made our work more difficult.

So, why a company should consume resources to create a social media honeypot? The reason is that even if the spearphishers do not study the psychological profile of a victim prior to an attack through social media they have knowledge of psychology and it is just a matter of time that they will exploit new capabilities for better results. As Mikko Hypponen of F-Secure said in his speech in IT College “fixing social problems is much much harder and that is why these problems will not go away any time soon”. So the public and private sector should take proactive measures such as social media honeypots to fight increasing social problems like phishing and spearphishing.

A different example could simulate the behavior of a malicious insider. We saw previously [12], [13] that there is a way to predict possible malicious insiders according to their YouTube and Twitter postings and according to their gender, emotional state and personality trait they belong into. A malicious insider [13] belongs to the

personality trait of Narcissism. A narcissist combines characteristics from 3 main personality traits, Extroversion, Neuroticism and Openness to experience. 90% of insider cases are connected with serious employment crises and 75% of them are males [13]. A big company which plans to lay off a number of people could use such social media honeypots of fake employees expressing in the social networks high neuroticism and anger for what is happening in the company. The fake employees will behave online shown signs of possible malicious insider. Bellow, is a table with a mixture of characteristics from the 3 personality traits that if combined can create a narcissist's fake profile which can later be used as a honeypot for attackers looking for malicious insiders.

	Malicious Insider	**Comments**
Personality trait	Narcissist	Combines characteristics from Extroversion mainly, Neuroticism and Openness.
Feel	Leadership, Authority, Exhibitionism	Glamorous self-promoting pictures
Friends	High	Very large numbers, Extrovert
Status updates	High	About themselves, Extroversion and Openness
Gender	Male	75% are Males
Topic of discussion	Against law enforcement, Authority, negative emotions, revenge syndrome	Neurotic
Behavior	Easy to befriended and be approached.	Extrovert's online behavior is closer to one with characteristics of Narcissism.

Table 4. Characteristics of malicious insider online behavior

	Likes	Friends	Groups	Tags	Status Updates	Photos	Language	Behavior	Results
High levels of									
Neuroticism	HIGH	MEDIUM	MEDIUM				I, me, mine, Acronyms, Anger expression, f***, depression, hate, alone, stressed, sick of, worse, annoyed, pissed, dead	Emotional unstable	An emotional unstable person seems an easy target for an attacker. **Reply to phishing emails.**
Extraversion	HIGH	HIGH	HIGH		HIGH		Express positive emotions, talkative, party, girls, can't wait, chillin', love you, night with, baby, weekend	Outgoing and friendly to people they do not know. Make friends easily.	Easier to be approached by an attacker. An extroverted person with a mixture of characteristics of high openness is similar to narcissism which is the characteristic of a **malicious insider**.
Openness	HIGH		HIGH		HIGH		Express, curiosity, imagination, adventurousness, new and strange ideas, express emotional sensitivity, tolerance and political liberalism,	Seeking new experiences, they like new and unusual ideas.	They might find idea of **malicious insider** exciting, because of tolerance they might discuss about it.
Agreeableness	LOW			HIGH			A, an, the, I, me, mine excited, amazing, wonderful, prayers, blessed, lord, awesome, merry Christmas, thanksgiving, Christ, thank you	Friendly and compassionate, behave in a cooperative way, trust other people and adapt to their needs.	**Opportunity to compromise and may be gullible. Great target for Phishers and Spearphishers.**
Conscientiousness	LOW		LOW			HIGH	Work, ready for, blessed, wonderful, long day, success, relaxing, thankful, vacation, workout, great day	Organized, reliable, consistent, seek achievements and pursue long-term goals.	Their large number of photos might give out personal information about what their routine is and which are the people they are connected to. That is a characteristic that could be used by an attacker.

	Likes	Friends	Groups	Tags	Status Updates	Photos	Language	Behavior	Results
Low levels of Neuroticism	LOW	MEDIUM	MEDIUM				Great, sports related, work out,beach,praised,blessed,God,week-end.	Calm and self-confident	Need more information to become a target.
Extraversion	LOW	MEDIUM	LOW		LOW		Anime, manga, pokemon, computer, windows, program, episode, internet	Not much interaction with outside world	Not easy to be approached.
Openness	LOW		LOW		LOW		Use of quotations, mate, talk, they, child, music, art, writing, dream, universe, soul		Not welcome the new and unusual ideas. Not a good target for an attacker.
Agreeableness	HIGH			MEDIUM			Use of bad language, words that correspond to hostile attitude and substance abuse	Cold and suspicious	Not a good target for an attacker. **In case of substance abuse that might be the key to exploit**.
Conscientiousness	LOW		HIGH			LOW	Words that reflect being disorganized, irresponsible, impractical, careless, f***, pokemon, youtube, facebook, bored, anime, gay	Easy-going, spontaneous and creative, more tolerant, less bound by rules	Low levels in this category give the impression of a mixture of high in Openness and Agreeableness which is a **good target for an attacker**.

Table 5. Personality traits and social network activity, language, behavior

8. The method

The more traffic and interaction the honeypot has with the intruders the more valuable is. So, in order to create a one way road to the high profile employee through the Personal Assistant (PA) there should be a pattern of how the high profile employee is connected with the world otherwise the honeypot will not be able to provide sufficient protection. The first step is that the company should not advertise company's emails in their website. At the contact section should be a form of communication. That way it would be more difficult for the phishers to collect the emails of the personnel. At the same time, the high profile employee has family, friends, university ex-classmates, colleagues and acquaintances. If he has social media accounts he should only allow people he personally knows well to be connected with him. At the same time he should activate all the relevant protection measures the social networks provide. For example, LinkedIn offers a service that if someone would like to connect with another person he needs to provide proofs (e.g. valid email address) that he knows that person. Any incoming emails from email addresses not included in the white list are rejected. Any friend requests from unknown people claiming to know or to be friends of friends are rejected. With that type of measures the high profile employee could create a white list or circle of trust. The rest of the people he communicates through his PA. A high profile employee might have more than one PAs. The PAs should be trained with anti-phishing embedded delivery method. After creating the above secure environment we deploy the fake employee as one of his PAs. In different working environments the fake employee might have multiple instances or different support. For instance a Ministry which has Minister, Deputy Ministers, Secretary General and many different Directors it is normal to have many fake personalities that will play the role of the victim. The roles the victim will have in the different environments will vary. For example, the army might have a victim which has characteristics of a malicious insider while a Ministry might have a victim which has the characteristics of a person who likes to talk a lot and share information from the work environment. All these will be decided according to the environment, the high profile employees and the assets we would like to protect.

In this project we deployed the honeypot in a small fake company. The Director of the company and the Personal Assistant (PA) are the main characters of the scenario. The fake company “NATON IT and Cyber Solutions” was referenced in our fake Facebook and LinkedIn profiles as place of work. There was no private mail server of the company, thus, a free webmail was used for that purpose. Mail.com was picked as email provider because no telephone numbers or other extra information was asked in order to register as users the two fake persons.

Their email was of a form of a corporate email name.surname@mail.com. The Director’s contacts were not available to anyone but the PA. The PA uploaded photos from the Internet that they were used for online advertisement. At some point it was offered by one female friend a photo to be used in the fake PA’s Facebook account. It was used but not as a profile picture. The reason was that we did not want an acquaintance of her to notice it and report the profile to Facebook. Additionally, we provided the opportunity to someone who is interested in proofs to find a picture of a real person in the profile.

When Facebook accounts were created the use of a mobile telephone number was needed to activate the accounts. Two separate mobile numbers were needed in total for the two fake profiles. In Estonia someone can obtain a mobile SIM card without registering his personal data together with the mobile number. The cost of each card was 1 euro.

During the project all friend requests to us were accepted. The PA’s profile initially was with no friends. Thus, the PA picked randomly friends from different geographical locations. The plan was to create an initial number of online friends and then upgrade the quality of the profile with online friends from Estonia such us professionals working in Estonia and TTU students. We made an effort, using the information from Table 5, the PA’s profile to reflect a mainly agreeable and extroverted behavior. When a user was asking from the fake PA to like another Facebook page the PA was doing that, showing signs of obedience. Extroverted behavior was difficult to succeed because there were no supporting characters to exchange activity on each other’s profiles or many real photos from participating in different real events. The behavior characteristics we tried to simulate are summarized in the following table.

Agreeable and Extroverted in Facebook	Level	Comments
Likes	Medium to Low	Combining a High of Extroversion and Low of Agreeableness.
Friends	High	There is no reference on the number of friends an agreeable person has, so, it is picked the extrovert characteristic that does not contradict with Agreeableness.
Groups	High	Same as in Friends.
Tags	High	Agreeable person are tagged in pictures a lot. Support profiles are needed to provide these numbers.
Status updates	Medium to Low	Given the fact the Agreeable people are careful not to give a bad impression to others they would avoid express themselves openly in public.
Photos	Medium	There is no reference in the research papers on the number of photos they upload but we assume that an extroverted behavior requires at least a medium number of photos
Gender	Female	Females have high score in Agreeableness
Language	A, an, the, I, me, mine excited, amazing, wonderful, prayers, blessed, lord, awesome, merry Christmas, thanksgiving, Christ, thank you, shopping, love you, my hair, happy, wishing, family, cute, mommy, her	A combination of Agreeable people and Females online

Agreeable and Extroverted in Facebook	Level	Comments
Topic	Relationships, proud or embarrassing things they did, good or bad times they had and things that make them feel great or crazy	
Age **Education**	24 (18+ Higher level education) Education, Business, Science	Less than 25 years old have the highest scores in Agreeableness
Behavior	Friendly and compassionate, behave in a cooperative way, trust other people and adapt to their needs	

Table 6. List of characteristics to simulate an Agreeable and Extroverted behavior online

The PA joined different Facebook groups, and mailing lists. The mailing lists varied from Clothing and Fashion to Reddit and Ruby programming language. We used this wide range of fake interest in order to advertise the email of the PA to different areas hoping that it will reach the lists of a phisher faster. At the end the PA registered her email in two adult content websites. We have noticed that the websites that are willing to share their email databases to other companies during registration they offer to the user the option "I do not want to share my information to 3rd parties". We didn't click on it, simulating a user's behavior that didn't notice that option. The goal of the aforementioned scenario was to receive phishing and spearphishing emails at PA's email address. We received only a few phishing emails of 4-1-9 Nigerian Scam type and one of them was using the last name of the PA in the subject section. That is the closest to a targeted attack that we received. More on the received emails will be discussed in the results section.

8.1. Deployment in an organization

At an organization before deploying the honeypot the high profile employees should create their circle of trust. Following the pattern we mentioned earlier the high profile employee, if he has Facebook and LinkedIn accounts he should allow only people who personally know well to be connected with him. At the same time he should activate all the relevant protection measures the social networks provide. Emails reaching his mailbox claiming to be from friends of friends he will not consider processing them. All kinds of communications from people he have not approved personally and they are outside his circle of trust are rejected and only the ones who are coming from the PA's side are counted as valid. A high profile employee might have more than one PA. The PA's should be trained with anti-phishing embedded delivery method. After creating the above secure environment we deploy the fake employee as one of his PA's.

The fake PA in a corporate environment should have everything the normal employees have. That means that she should have a corporate email of type name.surname@company.com, a telephone number in company's phone list, a job description, a job position, a CV, a working background and possibly some references in the company's website should exist. The last one will help the search engines create results for the fake PA and increase the impression that this is a real person. The method is presented in the following graph.

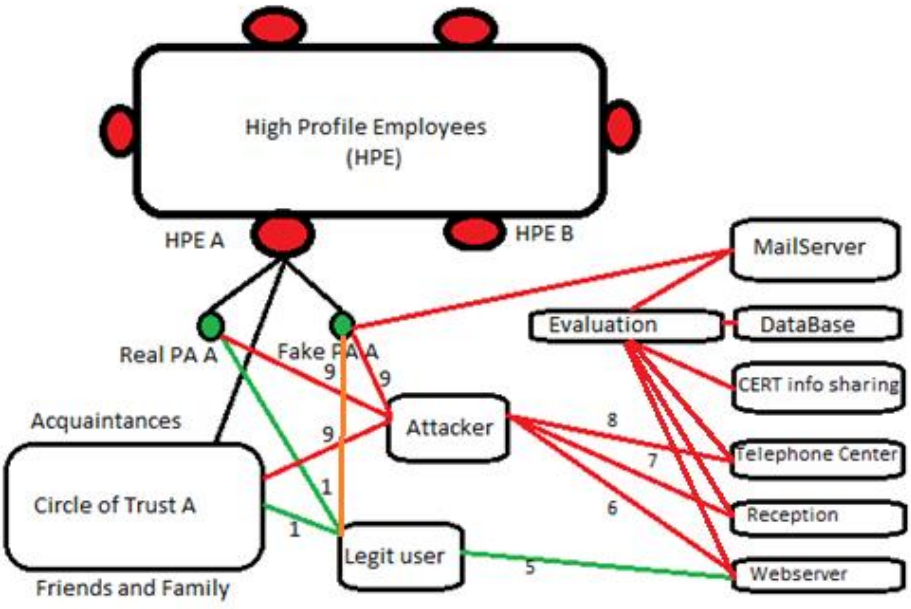

Figure 2. Method diagram

In the above scenario we have the high profile employee A (HPE A) which has 2 PA's, real PA A and Fake PA A. We have 1 attacker, 1 legit user a circle of trust which contains the friends, family and/or some of the acquaintances of HPE A. There are also a mail server of the company, a Database to store relevant information about the phishing attacks, an evaluation procedure and a channel of communication with the CERT for exchanging phishing and spearphishing information.

In the first option legitimate user (communication with green lines) who has previously contacted a person from the circle of trust of HPE A receives the information that in order to contact with HPE A he should contact Real PA A first. His email 1 first goes to one member of HPE A circle of trust and then, after he is notified whom to contact, he forwards the email 1 to Real PA A. Real PA A will decide if the email will be forwarded to HPE A or if she will process it herself according to HPE A directions.

In the second option an attacker (communication with red lines) which makes use of social engineering techniques tries to find a real email or other communication channel with HPE A. By social engineering, social and business environment of the HPE A, the attacker succeeds to collect the emails of the two PA's and some emails of HPE A's circle of trust. In this step he has two options, to contact all available persons and send email 9 to all of them or study them, possibly through social networks, and contact the appropriate person later. In the first option an email, among others, is sent to Fake PA A. The mail server receives email 9 and sends an alarm to the administrator which he evaluates the content of email 9, stores to the Database the available information extracted from the email such as FROM address, REPLY TO address, content of email body, persuation methods and/or special language characteristics if they exist. After a phishing or a spearphishing attack has been identified all emails received by the mail server from that FROM address or contain that REPLY TO address are rejected. Before the email to Fake PA, similar emails from the same phishing address might have reached other recipients in the organization. It is important the administrator to check if there have been any replies to the phishing address already. Outgoing emails to the phishing email address are dropped as well. At this point our honeypot works similar to Uri Rivner and Idan's Aharoni honeypot. Then the phishing information can be shared

with a security information sharing organization. One other option is the attacker to attempt to phish the targets through social networks like Facebook. In that case, the attacker might study the psychological profiles of their future victim to identify which one is the best choice. The Fake PA A is managed according to the findings of cyber-psychology gathered in Table 5 and the profile is revealing an agreeable and extroverted character. If the attacker at any point of communication through Facebook asks for a communication with HPE A the Fake PA A will respond "Send me an email in order to forward it to him". When email 9 is received from mail server an alarm to the administrator is generated as in previous case. Then the administrator evaluates the content of email 9, stores to the Database the available information extracted from the email such as FROM address, REPLY TO address, content of email body and/or special language characteristics if they exist. After a phishing or a spearphishing attack has been identified all emails received by the mail server from that FROM address or contain that REPLY TO address are rejected. Same as previously, all outgoing email to the phishing address are dropped as well. The phishing information then can be shared with a security information sharing organization. In case the legit user contacts the Fake PA A (orange line), he will be treated as attacker until is proven otherwise.

There is always the possibility that prior to an attack the attackers might want to verify that the person actually exists and they might telephone to the company to talk to her or pay a visit in the organization's premises. For that reason a Simple Event Correlator[19] is used to relate different events and create a better evaluation process. SEC will read logs of telephone calls to the fake PA, system notes from the reception desk that someone requests a meeting with the fake PA and the email activity to fake PA's mailbox. SEC will extract all important information from the above actions and if necessary will create alerts with a timestamp for the administrator and the data will be stored in the database. SEC will also create simple notices when someone visits (6) the company's website at the section with the information about the fake employee(s). The notice with the IP address accessed that webpage will be stored in the database and send to the administrator. All the aforementioned gathered information will be used to predict, prevent or help the investigations of an attack.

[19] SEC - Simple Event Correlator , https://simple-evcorr.github.io

9. Legal and Ethical considerations

Many organizations use honeypots to provide an extra level of security for their assets. In our case an organization would use fake employees with personality vulnerabilities, that will be advertised in social media such as Facebook and LinkedIn, as honeypot and we will explore if legal or ethical issues are raised by this action. Honeypots in general have their own documented legal issues as some security experts argue. At the same time our honeypot which interacts and deceives an attacker or a legitimate user has its own legal and ethical issues to be explored. The issues with the honeypots according security experts and Symantec [34] are entrapment, privacy and liability. The issues with our honeypot are online impersonation and deception in addition to the general honeypot issues.

The use of honeypots in security systems for some people is a controversial matter and has been widely discussed. The legal issue as it was mentioned before concern entrapment, privacy and liability. Entrapment according to various online dictionaries is when a person "is induced or persuaded by law enforcement officers or their agents to commit a crime that he had no previous intent to commit." [35] Privacy is about the information gathered from honeypots concerning personal private data with the ones interacting with the honeypot and liability is about who is responsible when, for example, an unlawful action starts from a compromised honeypot.

An organization that is using honeypots to protect its assets according to some security experts might not have the right to do so due to the legal and possible ethical issues raised. In order to defend our work with the proposed honeypot we will take the role of the devil's advocate in this debate. Our primary argument is that when it comes to security of State information or industry research or medical and banking data or any personal information that someone believes that it needs to be protected from disclosure, there are measures that need to be taken in order to increase security but at the same time there are some loses on other areas. A simple example is that in all organizations the computer users do not have administrator privileges to their computers. This provides an extra level of security because malicious or other

executables are not allowed to be installed but offers less freedom to the computer user and extra work to the support staff to install a new program every time it is needed. Each time, when this talk occurs the same argument is used, more security, less freedom in certain areas. The best solution is somewhere in the middle depending on what is in stake.

The first general honeypot issue is entrapment. Is it entrapment when monitoring the techniques of an attacker in a honeypot? According to the definition in order to have entrapment we need to have law enforcement officers or agents. A private organization or company is not law enforcement and does not have agents of law enforcement. The organization that is using honeypots to protect its assets is not asking from attackers to hack its systems in order to prosecute them. Even if the attackers compromise a honeypot that belongs to the Police, entrapment does not exist before someone commits a crime and it does not exist as an accusation. To monitor an attacker in a honeypot it means that first the attacker compromised illegally the honeypot and then he was monitored. Entrapment is an excuse used by the people who have been prosecuted for some crime in order to defend themselves and be released. The lawyers of the defendant are trying to prove that their client has been lured to commit a crime that would not do otherwise and it does not exist as an accusation before the crime is committed. Thus, we cannot accuse a honeypot system as illegal, trying to entrap attackers before they perform their attack.

In our case, the honeypot interacts with different people. Makes comments, “likes”, and creates status updates which are all false, having the purpose to deceive the people who have discovered them or the people who have been contacted by the honeypot. The difference of our honeypot with the general idea of the honeypots is that our honeypot is contacting different kind of people in order to create a profile of realism otherwise it cannot function properly. Our honeypot gives the impression of a vulnerable, agreeable person that could be exploited. If some people, legitimate users or attackers, think that it is a good opportunity to exploit that vulnerable person is it entrapment? By definition, no. If we raise it to the ethical level though, could it be an ethical entrapment? We would say no again. If we have our window open and a person passing by thinks “it is an opportunity to break-in”. Are we ethically wrong? Did we entrap that person? The

answer again should be no because we did not force that person to break-in but he decided by himself.

Entrapment seems from our point of view not to be an issue, legal or ethical, with our honeypot. Privacy issues though are a big chapter. There are issues about which country the honeypot operates in. If it is the US, the US Privacy Law applies. But is it the federal law or the state law? What if the honeypot is in one US state and the attacker in another? Most of the times, according to Symantec, it is the Federal Wiretap Act[20] that applies to these cases. If the honeypot is used to protect the organization it would fall under the Service Provider Protection exemption[21]. When a honeypot is used to gather information about the techniques of the attackers or who the threat is, then is not working directly to protect the organization and is less likely to fall under the above mentioned exemption.

In Europe there are different Directives dealing with data protection issues, e-Privacy Directive or Directive 2002/58/EC [14] and Data Protection Directive or Directive 95/46/EC [15]. The first Directive is focusing on "the protection of personal data and the protection of privacy in the sector of electronic communications" [36] while the second Directive focuses on "the protection of individuals with regard to the processing of personal data and on the free movement of such data".

We have the legal framework of data protection and we need to explore what data can be collected and under which circumstances. There are two categories of data, transactional data and content data. Transactional data are the data about how the connection between two points has been established. IP addresses, timestamps, protocols, MAC addresses, geo-location data, email headers and anything else that could be used to identify the persons that established the connection. Content data is the content exchanged between the two points of communication. For example, the main body of an email or the attachment of an email or the filename of the attachment, full network packets [36] for example VoIP packets capture that help to reconstruct the

[20] The Federal Wiretap Act was later amended by the Electronic Communications Privacy Act (ECPA), https://www.law.cornell.edu/uscode/text/18/part-I/chapter-119 and https://it.ojp.gov/PrivacyLiberty/authorities/statutes/1285#contentTop

[21] Exemptions in paragraphs (b) and (c), https://www.law.cornell.edu/uscode/text/18/2702

communication between the two parts. In our case the proposed honeypot interacts with different people, maybe even exchange messages which include personal information of the other side. The defense system of the organization is storing data, like IP addresses and timestamps of people visited specific parts of the company's website or accessed files like CV's of company's personnel. In special cases these files might contain data mining in order to reveal information about the computer system that accessed them.

All the above data can be collected by a honeypot but is it allowed to be processed? The e-Privacy Directive sets the legal ground for processing the above mentioned collected data. When someone is delivering a service, safeguard the security of the service and minimizing congestion, is allowed to process traffic and communication data [36]. In our case the honeypot has two parts. It is used in the organization to protect and mitigate the risk but it is also used to gather information about the attackers. For the first part we believe there are no legal issues according to e-Privacy Directive but for the second part it requires the consent of the employees in the company but we cannot have the consent of the attackers. At the same time the honeypot needs to be kept secret otherwise its role will be canceled. So the HR needs to inform the personnel for the use of security measures that apply in the organization and acquire their written consent for collecting personal data without revealing information about the type of the security systems or their nature.

A real example of such privacy issue happened during our project. We were interacting with different people and one of them revealed to us information, that someone could infer sexual orientation preferences. Such information especially if they come from inside the organization should not be processed or disclosed because they give no extra value to the initial goal of the honeypot. In this case the disclosure of such information is not only legal but ethical issue as well.

Another issue with honeypots is liability. There is a term called downstream liability [37] which deals with who is responsible when an attack occurred from a compromised honeypot or who is responsible when customer personal data have been stolen from a company's database etc. Is it the organization or the company responsible for the attack started from their compromised systems or the loss of data? Is it responsible the owner of a zombie pc for a DDOS attack? What if an attacker compromises a honeypot and

starts a streaming service from the honeypot? That is why the honeypot needs to be monitored closely and extra firewall rules should apply concerning, for example, the outgoing traffic of a honeypot. In our case, an attacker might find out that the specific fake Facebook and LinkedIn profiles are used as social honeypot. He might then change his strategy and attack on the Facebook and LinkedIn accounts. If the accounts are compromised he might gather information about the people the honeypot interacts with. The attacker might start creating status updates with insulting content for the company, the employees and the people the honeypot is associated with.

To understand the problems occurred by downstream liability we will use information found in a G.I.A.C paper "System Security Liability. Does it always float downstream?" [38]. GIAC is the Global Information Assurance Certification and has written a report analyzing the "potential threat posed by information systems negligence and liability suits to companies". Downstream liability has its roots to the legal theory of negligence. Negligence is when someone is omitting to do something which a reasonable person would do or do something which a reasonable person would not do. By our negligence some third party, company or person, might lose a type of property such as financial loss or intellectual property loss etc. Honeypots are used in order to protect information systems by attracting attackers and they are left intentionally with vulnerabilities. So, to prove liability there must be negligence, to prove negligence must be expectation of duty according to the GIAC report [38]. When the honeypot and the findings of the honeypot are closely monitored, evaluated and shared with other authorities like the national CERT to increase the organization and community's security then the administrator and the security department is meeting the expectations according to their duties. If the administrator and the security team have taken the minimum or more of the expected measures to secure the honeypot from creating illegal traffic then there is no negligence. Thus, we believe that when all of the above requirements are met concerning the functionality of the honeypot there is no entrapment, privacy issues or downstream liability.

Additional issues concerning our honeypot are online impersonation and deception. The online impersonation that takes place in this project is limited to some fake employees and is performed by creating fake social network accounts. Facebook forbids users to

have multiple accounts or use nicknames in order to avoid identification. At the same time the e-Privacy Directive does not allow the process of transactional or content information that identify the person behind a connection. So, could it be legal consequences because we created a few fake Facebook and LinkedIn accounts for a good cause? Without having a legal background to evaluate that, we believe no. What might happen is that Facebook or LinkedIn will delete our fake accounts if that is proven.

The other issue with our honeypot is deception. Yes, the honeypot is made to deceive the people is interacting with. The honeypot intends to attract attackers and deceive them while at the same time is in social networks and will be discovered by other users too. The honeypot might be discovered by company employees as well. It is possible an employee to read a website announcement referring to the fake employee and out of curiosity he might look for a way to contact the honeypot. For sure, the employee will be deceived but are we positive about the employee's intentions? He might be working for an attacker gathering information for company's employees. Should we uncover the honeypot in order not to deceive that person? We believe no. After evaluating the intentions of the employee who contacted the honeypot, the communication between the two parts should be minimized. On the other hand, the honeypot can be used for evaluating the online behavior of the employees. This is like walking on thin ice, because from one side are the interests of the organization and the other side are the personal data protected by the e-Privacy Directive. That part should be avoided unless the company has the written consent of the employee that clearly states that he accepts to be part of such a process.

In conclusion, we believe that honeypots cannot be accused for entrapment. Privacy issues exist but they are covered by the solutions provided by the US and EU Law. Liability issues can be overcome when the companies take appropriate measures to protect their valuable information. So, from the above we can infer that there are restrictions in collecting, storing and processing personal data information and there are risks concerning the use of honeypots but there is also ground to use the honeypots in such a way that will not abuse the information collected and meet the final goal which is to increase the security for the organization and the Internet community in general.

10. Results

For about 4 months we deployed a social media honeypot conducting an experiment to detect spearphishing. The honeypot would protect the high profile employee of a fake company named “NATON IT and CYBER SOLUTIONS”. We created for the high profile employee and his personal assistant social network accounts in Facebook and LinkedIn and the name of the fake company was advertised there. We wanted to test how fake social media network profiles can be used to detect spearphishing and for that reason we build a method to create such profiles that reflect certain personality traits.

In the beginning, the fake Facebook profiles had no friends and no history. Our initial goal was to add a number of friends that would reflect realism on the profile of the personal assistant. We did not want to do the same with the company’s director Facebook profile because according to our scenario he has a silent behavior in social media. We assumed that people who have joined and follow the same Facebook groups would be more open to accept our friend requests. For that reason we joined Music, Fashion, Tallinn expatriates and News groups and we start sending friend requests to their members. At a second stage we would enrich our friend base with Estonian people since our fake company is based in Tallinn.

We sent out 50 friend requests during these 4 months of the experiment and 30 of them they were accepted. We have received 13 friend requests and accepted all of them. Some of these profiles were deleted by their owners (or Facebook) during this period but we managed to have a friend base of around 40 people most of the time. The profiles that were deleted belonged to males and females, they had been created recently (within the current year) and they had a huge friend base of a few thousand of friends. They claimed to be from Estonia, Arabic, Asian and Latin American countries. In LinkedIn we have 9 connections plus 1 interesting case that will be analyzed. The overall friend base we created during this project was 43 people in Facebook and 10 people in LinkedIn and below is presented a demographic of them in two separate tables.

	Friend requests							
	Friend requests sent				Friend requests received			
	Accepted		Declined		Accepted		Declined	
	Males	Females	Males	Females	Males	Females	Males	Decline
Estonian Public Sector	0	0	1	1	0	0	0	0
Estonian Private Sector	0	2	0	9	0	0	0	0
TTU Employees	0	0	0	2	0	0	0	0
TTU Students Estonians	0	0	4	2	0	0	0	0
TTU Students non-Estonians	4	1	1	0	0	0	0	0
Non-Estonians	9	14	0	0	7	6	0	0
Total	13	17	6	14	7	6	0	0

Table 7. Personal Assistant's Facebook friend requests (Sent/Received)

The Facebook friends were 20 males and 23 females. From the 30 people which accepted our friend request 13 were males and 17 were females. From the 13 people we accepted their friend request, 7 were males and 6 were females. 20 of our friend requests were not accepted.14 of them targeted females and 6 of them targeted males. The female targets are, 1 working for Estonian public sector, 2 TTU employees, 2 TTU students and 9 females working at the private sector. The males are 1 Estonian public sector employee and TTU student, 4 TTU Estonian students and 1 TTU non-Estonian student. Most of the people that were picked to befriended are of young age because young people have higher levels of Agreeableness thus, are easier targets. The literature suggests that females are more susceptible to phishing than males but we discovered during this experiment that Estonian females do not accept easily a friend request coming from an unknown female which is a good first defense before a phishing attempt.

In LinkedIn we did not try to create a big friend base, it was used as an alternative for our Facebook friends to check our existence. And in fact, most of our connections are Facebook friends which they searched the PA's existence in LinkedIn. After we were notified by email that someone has checked the fake LinkedIn profile we sent a friend request which was accepted rather quickly from the other side except one request to a TTU lecturer. Following is a table containing some basic information about the people connected to us in LinkedIn. The 1 non-Estonian male is a former CCDCOE researcher and currently University lecturer abroad.

	Friend requests	
	Accepted	
	Males	Females
Estonian Public Sector	0	0
Estonian Private Sector	2	2
TTU Employees	0	0
TTU Students Estonians	0	1
TTU Students non-Estonians	2	1
Non-Estonians	1	1
Total	5	5

Table 8. Personal Assistant's LinkedIn connections demographic.

The one Non-Estonian female is an interesting case. Two days after the fake profile was created we found a Bloomberg article on a company "Naton Technology Group Co. Ltd"[22]. If someone reads the article is obvious that there is no information on the

Figure 3. Possibly real profile

22 Naton Technology Group Co. Ltd on Bloomberg, http://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=105765736

company but a Beijing address. It was an opportunity to share that article to our online friends because we assume that most of the people just read the headlines and not the content. So, we claimed that the NATON Tallinn based company, our fake company, is a branch of the main company based in Beijing. Three days after, a non-Estonian female requested a friendship on LinkedIn. Her name is Deborah Mullet claiming to be a Glasgow Lloyd's Bank senior accountant. We accepted her friend request and at the same time we made a research on her in Facebook. We discovered 3 profiles with the same name. One was almost empty with a few Latin American graffiti photos. The other two was one looking real of a middle-aged woman working in the medical section and the second one was using the previous profile picture claiming to be an assistant manager at Lloyd's TSB in Glasgow UK.

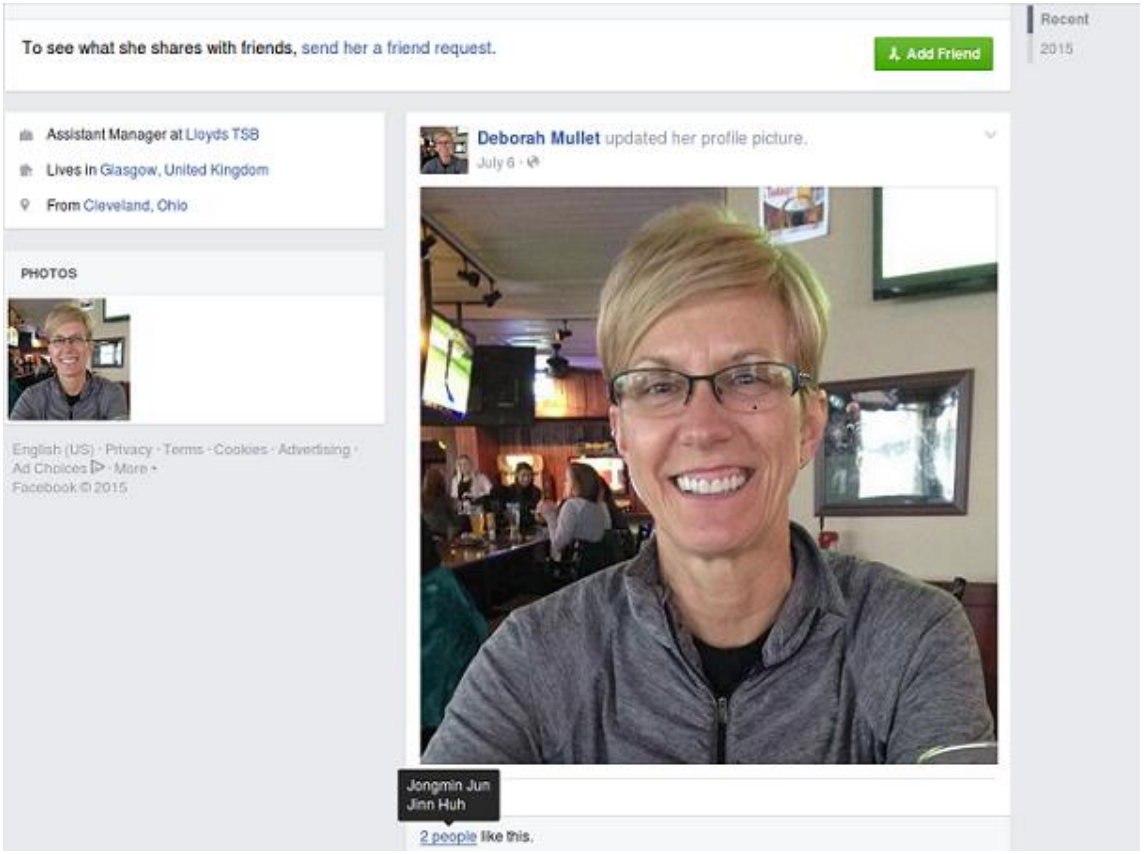

Figure 4. Possibly fake profile

From the above two images we notice that the real profile contains friends, interaction with the friends, pictures with family members and history. The fake profile has just one picture of the "owner", was created within the current year and has no history. The

interesting part is that the profile picture has been liked by two people possibly of Chinese origin, according to their names. So, some person two days after we shared a Bloomberg article on our Facebook profile about a Chinese company with very similar name to our fake company, contact us in LinkedIn. It seems that the persons we have as friends in Facebook are somehow associated with the person behind the fake Deborah

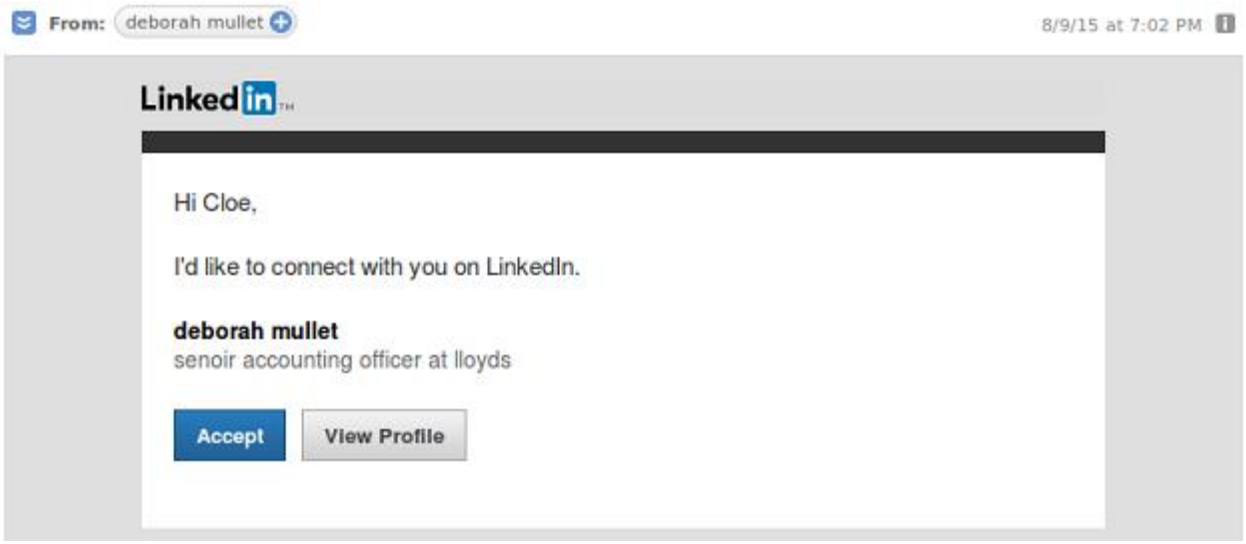

Figure 5. Deborah Mullet LinkedIn connect request

Mullet profile. The contact started with the email notification from LinkedIn which is shown below. One thing we notice in the email is the typographic mistake on the title of the employee. It is written “senoir” instead of “senior”. During the four months of our experiment this LinkedIn account has changed many times how the name is displayed (Deb or Deborah) or the country of origin. The next image shows the results of a search in LinkedIn for a Deborah Mullet. We notice our connection which is now changed to Deb Mullet from Nigeria. They have also corrected their typographic mistake from “senoir” to “Senior”. We notice also that there are two more Deborah Mullet accounts concerning Senior Accounting officer at Lloyd’s Bank in Portugal and a Senior Accountant at Lloyd’s Bank in San Francisco. In the middle of the results list we notice a Deb Mullet, nurse at a hospital in Cleveland, Ohio. The “real” Deborah Mullet is from Cleveland, Ohio, working at the medical section as she posted in her Facebook page. As we mentioned earlier there is one more Deborah Mullet Facebook account with Latin American characteristics. That profile completes an interesting square. We have some Latin American friends in Facebook and we have a Latin American Deborah Mullet Facebook profile. We have found a fake Facebook profile about a Deborah Mullet with

possible Chinese friends. We have found a Chinese company with a very similar name to our fake company with no activity and no other information about it and the square closes with the change in the country of origin in the Deborah Mullet LinkedIn account to Nigeria. We cannot be sure if they contact us in the first place because of the way we managed our fake Facebook profile or if we managed to attract cyber criminals in such a short time but there is a case of identity theft and a scam around the name of Deborah Mullet with a reference to countries or regions like China, Nigeria and Latin America which they have high activity in the eCrime area.

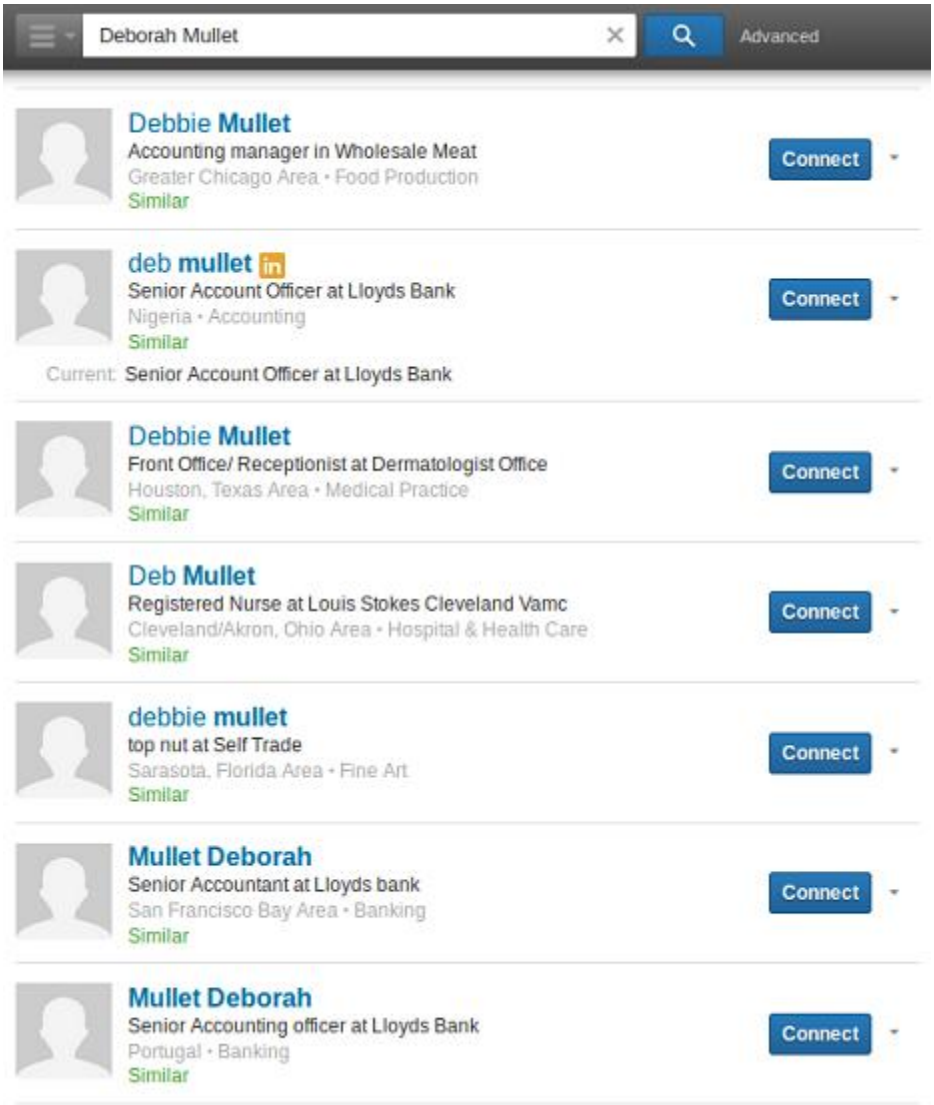

Figure 6. Deborah Mullet LinkedIn search results

A second interesting case has the name “Jana Clark”. Jana Clark’s profile was created

within the current year and had a few thousands of friends. Couple of days after we became friends and even though our profile is of a young woman's profile, we were invited to join a group named "Naughty Zone". In the group's Facebook page Jana was posting adult content pictures advertising a link for watching live what she was doing. The link was http://bit.ly/letscumtogether. We checked it in VirusTotal[23] and it was found as malicious from 1 (Blueliv) out of 65 URL Scanners. The link after the redirections is the http://www.ahookuptonight.net/ and it was using the IP address 69.58.188.39. The fact that someone is offering for free a product that has a value and there is out there a group of people that would pay for it, is making us very suspicious, even though, the majority of the URL Scanners found the website as "Clean". Few days after, Jana Clark's profile was deleted and a new girl took over with a new name. The Naughty Zone group seemed to have had 5 administrators which they kept the group alive for few more days and shows and then both Facebook profile and group were deleted.

Next are two 419 Fraud attempts. The first one did not mention any specific

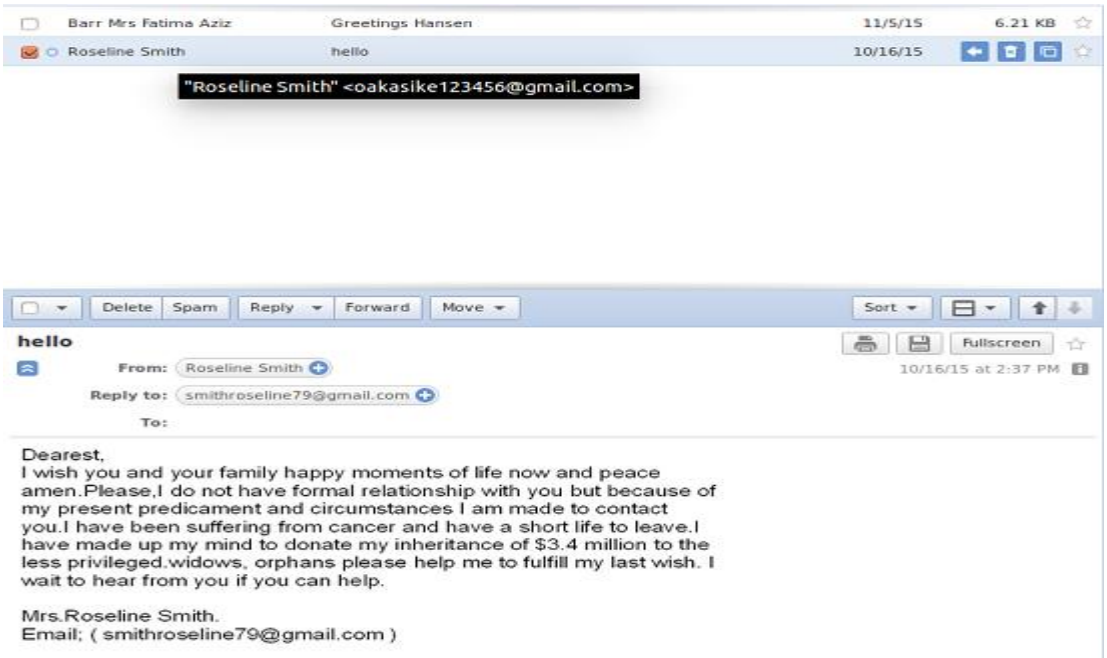

Figure 7. Phishing email

[23] www.virustotal.com

country but the second one mentioned Dubai. The first email was a photo with the text on a transparent background. Typical method used by phishers to bypass email content filtering. In the next image we see the email and the different email addresses used. We notice here that the REPLY TO address is different from the FROM address. We see that Gmail is used for this phishing email and the email address that came from is "oakasike123456@gmail.com". The way they refer to us is as "Dearest". They use politeness, a disease like cancer and a noble cause such as donating a big amount of money to underprivileged, widows and orphans to create emotion in order to make the victim to reply to the email. The next case is similar to the previous one but they do not use a picture to display their message on but text. They use United Arab Emirates as country of origin for their story. This case uses urgency, politeness and a big amount of money as bate. The mailer used for the phishing email is Gmail again and the name on the REPLY TO address is totally different than who they say they are. Someone with an email as Gerrard Bradford claims to be Mrs Fatima Aziz. The difference with the previous case is that they use our fake profile's last name Hansen to address to the PA.

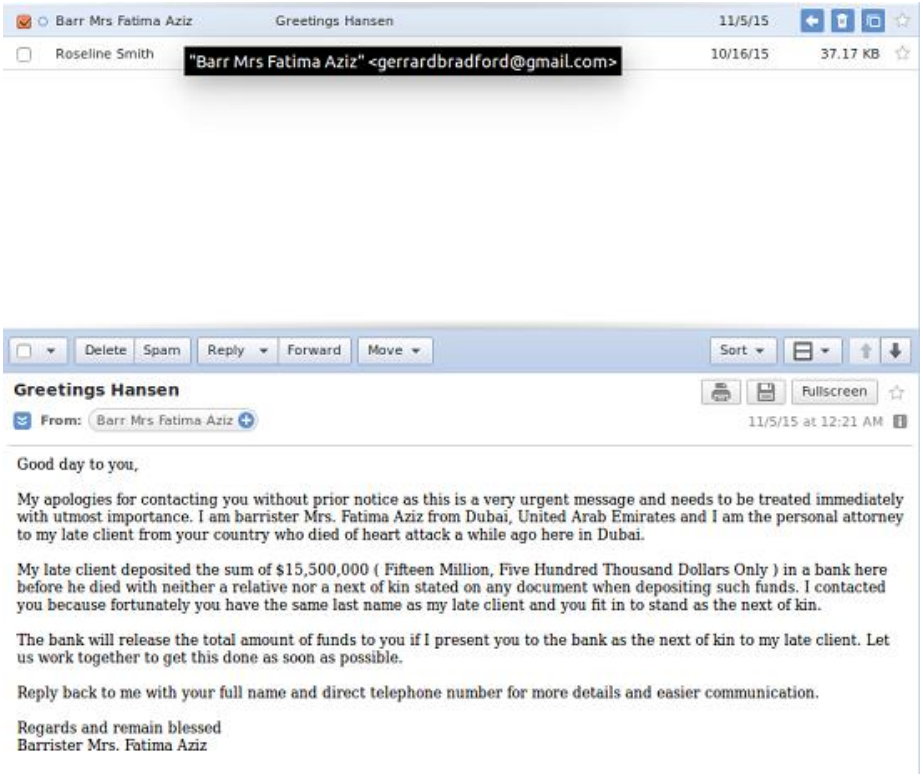

Figure 8. Phishing email II

In this section we saw some different methods used to exploit weaknesses in the victim's personality but we did not succeed to receive a spearphishing email. The closest was an email of 419 Fraud which used the last name of the PA in the subject area. The body of the email did not make any specific references to the PA or to information harvested by the social media.

During this project we contacted a few TTU students and one of them agreed to accept us on a closed TTU group in Facebook. Tracking the conversations in this group it was not difficult to understand that the majority of the students are anxious about finding a job in their area of studies. Using exactly that as a subject we managed to have extensive discussions with the TTU students that accepted our friend request. We were asked out to dinner and drinks by male and female friends. The ground seemed fertile to exploit their weakness but we did not. Another weakness that was obvious after a few lines of conversation was that some of them they were very open to having social relationships. As we saw previously in Robin Sage case, physical attraction seems to be stronger than the knowledge we have about cyber crime. During CyCon[24] 2015 a high profile employee of a NATO country mentioned to us that the diplomatic personnel that comes from the industry is more aware on cyber threats and has more professional behavior when it comes to handling classified information than the personnel which has no previous experience in the private sector. Studies have shown also that the diplomatic and military authorities are in the top of the target list. Our suggestion is that high profile employees and people having job positions of responsibility they need to have anti-phishing training with embedded delivery method. They need to feel all the negative emotions described in [22] in order to be more prepared. The humiliation, anger and the denial someone is feeling when hacked or phished it seems to be the best training for a better online behavior.

[24] CyCon is a Cyber Defense Conference held in Tallinn annually since 2009

11. Summary

In this document we describe a method to detect spearphishing with the use of social media honeypot. Our motive is that there are plenty of reports highlighting and warning that social media features are used by phishers and spearphishers but it does not seem to be much work in the literature that studies this phenomenon. We know from the relevant research that he attacker's goal is to affect the emotional stability of the victim and make him take decisions that he would not take otherwise. For that purpose they use persuasive methods such as Authority, Politeness, Urgency and triggers such as Alert, warning, attention, account verification and invalid login attempts. Other ways to lower the defenses of the other side are physical attraction like in Robin Sage case and impersonation like in CIA's Director Email hack. Our project is based on two hypotheses. First that the spearphishers study the psychological profile of the victim prior to an attack in order to decide which psychological technique will give the best result. Our second hypothesis is that if we create a fake person which carries the personality characteristics the attackers look for, then they will choose to approach that person. In order to give to the attackers what we assume that they want we created a fake company and a couple of fake employees that will play the role of a social honeypot. The one fake employee would be the Director of the company and the other the personal assistant. The Director is silent and the PA has a more loud presence in social media. She tries to socialize and advertise the characteristics of her personality.

The theory of Big-5 or the five traits of personality analyze the human personality in 5 main traits, Openness to Experience, Extroversion, Conscientiousness, Agreeableness and Neuroticism. People with high levels of openness are open to new or unusual ideas while extroverted people are easy to connect with. Persons with high scores in conscientiousness are following the rules while agreeable persons tend to believe that people have mostly good intentions. Neurotic people basically express negative feelings and they are not social. According to some studies discussed previously, a combination of characteristics found in neuroticism and narcissism describe a malicious insider. Phishers and spearphishers would prefer their attacks to interact with agreeable and

extroverted people. The reason is because they can be approached easier and they believe the good intentions of the other side. Young people have high levels of agreeableness thus, they are easier targets. The education background, gender and/or anti-phishing training play a big role too.

Combining the above knowledge we create a social media honeypot using social networks like Facebook and LinkedIn to present and advertise the character of the fake PA. The Director is silent; he is not exposed online and is difficult to directly contact him. Thus, there is only one way through the PA to reach the Director. The fake PA can be useful only if the Director creates a circle of trust and is unreachable otherwise. If the trusted people are white listed then anyone who tries to contact him will have to be filtered by the PA. Anyone contacting the fake PA generates an alert to the system for a possible phishing attempt. All the emails reaching fake PA's mailbox are evaluated and their interesting parts are stored in a database (FROM, REPLY TO, IP addresses, special wording in the email body etc). If the security team of the organization decides that there are evidences of a phishing attack then all emails containing the FROM address or the REPLY TO address are dropped before they reach the recipient's mailbox.

We managed our fake Facebook profile of the PA for almost 4 months according to the findings of cyber-psychology and the various studies dealing with the victim's personality and online behavior. We received over 1500 emails during that time but we did not receive any spearphishing emails. The most interesting result came from LinkedIn. A person claiming to be someone else, having a job at a popular Bank made a friend request on LinkedIn. That person seemed to have several Facebook and LinkedIn accounts pointing to a possible scam. We also received an email with the PA's last name on the subject area but there were no signs of information harvested from our honeypot. A Facebook contact invited us to join a Facebook group which was promoting live adult content free of charge on a URL which seemed suspicious for phishing.

Online users face everyday this kind of threats and they need to be protected at their working environment or when at home. As it was mentioned before, the attackers are trying to affect the victim's emotional state and make them take decisions they would

not take otherwise. That means that we cannot rely totally on the user's reaction. The organizations or the ISPs need to take action and try to protect their users. The companies have interest to protect their assets by protecting their users from being phished or hacked. The ISP's by protecting their clients increase their credibility and possibly their customer base. To do that, sometimes security experts need to walk on a thin line between what is allowed and what is not by the national legislation, the legislation of a greater community like the EU or the international law. It is a wide discussion about how someone can collect evidences about the actions of cyber criminals but the legislation is giving the opportunities to security experts to protect the organization's assets and their users without turning the protection mechanism into a spying procedure against their own employees. It is important to understand that as the companies have every right to protect their valuable assets at the same time the employees or the citizens have every right to protect their privacy, even though, the majority of them is giving out for free most of their personal information in social networks.

An extension of our project would be to deploy a social media honeypot on a real company or a public organization and study the results of the proposed method. A real company has real threats than a fake company like ours. So, we believe that the number of attacks or suspicious contacts would be much larger. We assume that the companies would have an interest to deploy social honeypots to collect data of possible attacks and share with other companies their results. That way, they will not have to wait until the direct contact with the attackers but they can be more effective when they are using data collected by other companies. That cooperation in private sector can create a security shield for them and if the data are shared with the national CERT then that shield can be extended in the public sector as well. The public sector has high profile employees that are in the top of the target list. Politicians, Diplomats, Military personnel are handling classified information every day and there are many attackers that would try to harvest these information. Thus, the public sector has big interest to deploy social media honeypots next to its high profile employees. The goal of sharing this information between the private and public sector would result to a Public Private sector Partnership on that field.

As the technology creates new ways, new business and study opportunities for the people, the threats do not seem to be eliminated but they are growing bigger every day. While the machines and the networks become more stable the human aspect becomes the weakest link in the security chain. The more the people are interacting with the cyber space and they have access to their personal information, money or intellectual property through the network the online threats will increase. The online history is still very young and most of the people have not learned their lesson yet. After a period of innocence that we experienced the previous decade the people need to mature their thoughts and awareness on cyberspace for the next day to find them in a more secure online environment.

References:

1. Darwish, A., El Zarka, A., Aloul, F. Towards Understanding Phishing Victim's Profile, 2013

2. Kim Zetter, Teen Who Hacked CIA Director's Email Tells How He Did It, 2015, (http://www.wired.com/2015/10/hacker-who-broke-into-cia-director-john-brennan-email-tells-how-he-did-it/) (31.12.2015)

3. Thomas, R., Getting in bed with Robin Sage, Las Vegas: BlackHat USA, 2010

4. Brandon, A., Wilson, H., A Study of Social Engineering in online Frauds, 2013

5. Prateek, D., Anand K., Ponnurangam K., Analyzing Social and Stylometric Features to Identify Spear phishing Emails, 2014

6. Lance, S., (2002), Honeypots: Tracking Hackers, 61-62, [Online] (http://www.it-docs.net/ddata/792.pdf) (31-12-2015)

7. Van, N., Attribution of Spear Phising Attacks: A Literature Survey,2013

8. Blumer, T., Doering, N. Are we the same online? The expression of the five factor personality traits on the computer and the Internet, 2012, [E-journal] (http://www.cyberpsychology.eu/view.php?cisloclanku=2012121201) (31.12.2015)

9. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., Stillwell, D., Personality and Patterns of Facebook Usage, 2012

10. Schwartz, A., Eichstaedt, J., Kern, M., Dziurzynski, L., Ramones, S., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M., Ungar, L., Personality, Gender and Age in the Language of Social media: The Open-Vocabulary Approach, 2013

11. World Well-Being Project, University of Pennsylvania, The Language of Personality, Word Clouds, [Online] (http://www.wwbp.org/personality_wc.html) (31.12.2015)

12. Kandias, M., Stavrou, V., Bozovic, N., Mitrou, L., Gritzalis, D., Can we trust this user? Predicting insider's attitude via YouTube usage profiling, 2013

13. Kandias, M., Galbogini, K., Mitrou, L., Gritzalis, D., Insiders Trapped in the Mirror Reveal Themselves in Social Media, 2013

14. Directive 2002/58/EC of the European Parliament and of the Council, 2002, EUR-Lex-32002L0058 - EN, [Online] (http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32002L0058) (31.12.2015)

15. Directive 95/46/EC of the European Parliament and of the Council, 1995, EUR-Lex-31995L0046 – EN, [Online] (http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046) (31.12.2015)

16. Koch, R., Golling, M., Dreo, G., Attracting Sophisticated Attacks to Secure Systems: A new Honeypot Architecture, 2013

17. Uri, R., Aharoni, I., Protecting Electronic Assets Using False Profiles in Social Networks : US8,856,928B1, 2014, [Online] (http://www.google.com/patents/US8856928) (31.12.2015)

18. Hill, K., I created a fake business and bought it an amazing online reputation, 2015, (http://fusion.net/story/191773/i-created-a-fake-business-and-fooled-thousands-of-people-into-thinking-it-was-real/) (31.12.2015)

19. Rathgeb, E., Hoffstadt, D., The Email Honeypot System – Concept, Implementation and Field Test Results, 2008

20. Li, S., Schmitz, R.,A Novel Anti-Phishing Framework Based on Honeypots, 2009

21. Chauhan, S., Shiwani, S., A Honeypots Based Antiphishing Framework, 2014

22. Jagatic, T., Johnson, N., Jakobsson, M., Menczer, F., Social Phishing, 2007

23. Abbasi, A., Zahedi, F., Chen, Y., Impact of Anti-Phishing Tool Performance on Attack Success Rates, 2012

24. Khoji, M., Iraqi, Y., Jones, A., Mitigation of Spear Phishing Attacks: A Content based Authorship Identification Framework, 2011

25. Evans, D., Gosling, S., Carroll, A., What Elements of an Online Social Networking Profile Predict Target - Rater Agreement in Personality Impressions?, 2008

26. Salusky, W., Danford, R., An Ever Changing Enemy, 2007, [Online] (https://www.honeynet.org/book/export/html/130) (31.12.2015)

27. Number of monthly active Facebook users worldwide as of 3rd quarter 2015, 2015, [Online] (http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/) (31.12.2015)

28. Sarah Frier, Facebook's Second-Quarter Revenue, Profit Tops Estimates, 2014, [Online] (http://www.bloomberg.com/news/articles/2014-07-23/facebook-posts-second-quarter-revenue-profit-topping-estimates) (31.12.2015)

29. Haddadi, H., Hui, P., To Add or not to Add: Privacy and Social Honeypots, 2010

30. Kosinski, M., Matz, S., Gosling, S., Popov, V., Stillwell, D., Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines, 2015

31. Buffardi, L., Campbell, K., Narcissism and Social Networking Web Sites, 2008

32. Marshall, T., Lefringhausen, K., Ferenczi, N., The Big Five, self-esteem, and narcissism as predictors of the topics people write about in Facebook status updates, 2015

33. Kleinberg, J., The Small-World Phenomenon: An Algorithmic Perspective, [Online] (https://www.cs.cornell.edu/home/kleinber/swn.d/swn.html) (31.12.2015)

34. Spitzner, L., Honeypots: Are They Illegal?, 2010, [Online] (http://www.symantec.com/connect/articles/honeypots-are-they-illegal) (31.12.2015)

35. The Lectric Law Library, [Online] (http://www.lectlaw.com/def/e024.htm) (31.12.2015)

36. Sokol, P., Husak, M., Liptak, F., Deploying Honeypots and Honeynets: Issue of Privacy, 2015

37. Mack, S., What is downstream liability? [Online] (http://smallbusiness.chron.com/downstream-liability-65026.html) (31.12.2015)

38. Eardley, G., System Security Liability Does it always float downstream? , 2004, [Online] (https://www.giac.org/paper/gsec/4126/system-security-liability/106532) (31.12.2015)