1 BACKGROUND & LITERATURE REVIEW

In order to meet our goal, necessary knowledge on phishing in general is required. This chapter introduces an understanding of phishing, an exploration of its cost, an overview of human factor in phishing, the overview of its modus operandi, a brief review on types of phishing, an understanding of bad neighborhoods on phishing and its general current countermeasures.

1.1 WHAT IS PHISHING?

While the Internet has brought convenience to many people for exchanging information, it also provides opportunities to carry malicious behavior such as online fraud on massive scale with a little cost to the attackers. The attackers can manipulate the Internet users instead of computer system (hardware or software) that significantly increase the barriers of technological crime impact. Such human centered attacks could be done by social engineering. Acoording to Jakobsson, et al. phishing is a form of social engineering that aims to retrieve credential from online users by mimicking trustworthy and legitimate institutions [**?**]. Phishing has a similar basic principle as 'fishing' in the physical world. Instead of fish, online users are lured by authentic looking communication and hooked by authentic looking websites. Not only that, online users also may be lured by respond to a phishing email, either replying or clicking an obfuscated link within its content. There are diverse definitions of phishing in our literature reviews, therefore, we would like to discuss about its universal definition in later section. However, one of phishing definitions accoding to Oxford dictionary:

> "A fraudulent practice of sending emails purporting to be from reputable companies in order to induce individuals to reveal personal information, such as passwords and credit card numbers, online" [**?**].

Several studies suggest that phishing is form of online criminal activity by using social engineering techniques [**?**][**?**][**?**][**?**]. An individual or a group who uses this technique is called *Phisher(s)*. After successfully gaining a sensitive information from the victim, phishers use this information to access victim's financial accounts or committing credit card frauds. The technique of phishing may vary, but the most common technique of phishing attacks done by using fraudulent emails and websites [**?**]. A fraudulent website is designed in such a way that it may be identical to its legitimate target. While it may be true, phishing website also could be completely different with its target as there is no level of identicalness. However, Preliminary analysis on what changed or added in phishing website would be conducted in the later section.

1.1.1 *The History*

The first time the term "phishing" was published by the **AOL!** (**AOL!**) UseNet Newsgroup on January 2, 1996 and was started to expand in 2004 [**?**]. Since then, we considered phishing development in cyberspace has been flourishing by phishers to make profit. Total losses due to phishing in 2004 reached more than U.S. $ 2 billion, it was involving more than 15,000 sites that become victims [**?**]. We will try to discuss about direct and indirect cost at present days in the later section. Evidently, Jakobsson, et al. [**?**] mentioned that in the early years of 90's (according to [**?**] it was around 1995) many hackers would create bogus AOL user accounts with automatically generated fraudulent credit card information. Their intention to give this fake credit card information was to simply pass the validity tests performed by AOL. By the time the tests were passed, AOL was thinking that these accounts were legitimate and resulted to activate them. Consequently, these hackers could freely access AOL resources until AOL tried to actually bill the credit card. AOL realized that these accounts were using invalid billing information, thus deactivated the account.

While creating false AOL user accounts with fake credit card information was not exactly phishing attacks, but AOL's effort to counter against the attacks was leading to development of phishing. This countermeasure includes directly verifying the legitimacy of credit card information and the associated billing identity, forced hackers to pursue alternative way [**?**]. Hackers were masquerading as AOL's employees asking to other users for credit card information through AOL instant messenger and email system [**?**]. Jakobsson et al. suggest that phishing attacks were originating from this incident [**?**]. Since such attack has not been done before, many of users have been victimized by then. Eventually, AOL enforced warning system to the most of its customers to be vigilant when it comes to sensitive information [**?**]. At the present day, phishing attacks might not only being motivated by financial gain but also political reason, and they have been emerging not only aim to AOL users, but also any online users. Consequently, large number of legitimate institutions such as PayPal and eBay are being spoofed.

1.1.2 *The universal definition*

Before we begin to understand deeper about how and why phishing attack works, we will briefly explore common phishing definition. Currently, there is no consensus definition, since almost in every research papers, academic textbook or journals has its own definition of phishing [? ? ? ? ? ? ?]. Phishing is also constantly evolving, so it might be very challenging to define its universal terminology. There is not so much study that specifically addresses the standard of phishing definition. However, one research conducted by Lastdrager [?] addressed to achieve consensual definition of phishing. Before we comply with one consensual phishing terminology, we will take a look at various phishing definitions from other sources:

> *"Phishing is the act of sending a forged e-mail (using a bulk mailer) to a recipient, falsely mimicking a legitimate establishment in an attempt to scam the recipient into divulging private information such as credit card numbers or bank account passwords"* [?]
>
> *"Phishing is a form of Internet scam in which the attackers try to trick consumers into divulging sensitive personal information. The techniques usually involve fraudulent E-mail and web sites that impersonate both legitimate E-mail and web sites"* [?]
>
> *"Phishing is an attack in which victims are lured by official looking email to a fraudulent website that appears to be that of a legitimate service provider"* [?]
>
> *"In phishing, an automated form of social engineering, criminals use the internet to fraudulently extract sensitive information from businesses and individuals, often by impersonating legitimate web sites"* [?]

It is noteworthy that the definition described by James, et al, Tally, et al, and Clayton, et al. [? ? ?] specifies that the phishers only use email as a communication channel to trick potential victims. While it might be true because using email would greatly cost effective, but we believe that phishing is not only characterized by one particular technological mean, as phishers can also use any other electronic communication to trick potential victims (i.e private message on online social network). This definition is also similar to dictionary libraries [? ? ?] that mention email as a medium communication between phishers and users.

We believe that standard definition of phishing should be applicable in most of phishing concept that are presently defined. Consequently, the high level of abstraction and is required to build common definition on phishing. We also argued that the definition of phishing should not focus on the technology being used but rather on the methodology how the deception being conducted, an "act" if you will. Therefore, We follow the definition of phishing by Lastdrager [?] which stated:

> *"Phishing is a scalable act of deception whereby impersonation is used to obtain information from a target"*

According to Lastdrager [?], to achieve this universal definition, a systematic review of literature up to August 2013 was conducted along with manual peer review, which resulted in 113 distinct definitions to be analyzed. We thereby agree with Lastdrager [?] that this definition addresses all the essential elements of phishing and we will adopt it as universally accepted terminology throughout our research.

1.2 THE COSTS OF PHISHING ATTACKS

It is a challenging task to find a real cost from phishing attacks in term of money or direct cost. This due to financial damage for bank is only known by banks and most institutions do not share this information with the public. Evidently, Jakobsson et al. argue that phishing economy is consistent with black market economy and does not advertise its successes [?]. On this section, a brief explanation of direct and indirect cost on phishing attack will be illustrated based on literature reviews.

According to Jakobsson et al., direct cost is depicted by the value of money or goods that are directly stolen through phishing attack [?]. While indirect cost is the costs that do not represent the money nor goods that are actually stolen, but it is the costs has to be paid by the people who handle these attacks [?] (i.e. time, money and resources spent to reset people password).

As we mentioned earlier, the difficulty of assessing the damage on phishing attacks is caused by banks and institutions that keep this information themselves and the unwillingness of many users to share to acknowledge that they have been victimized by phishing attacks. This happens because of fear of humiliation, financial loses, or legal liability [?]. Evidently, studies estimate the damage ranging from $61 million [?] to $3 billion per year [?] of direct losses to victims in the US only [?][?]. In addition, the Gartner Group claimed to estimate of $1.2 billion direct losses of phishing attack to US banks and credit card companies for the year 2004 [?]. By the 2007, it escalated to more than $3 billion loss [?]. The estimation also performed by TRUSTe and Ponemon Institute that stated the cost of phishing attack was up to $500 millions losses in the US for the same year [1]. Recently, RSA FraudAction gives monthly reports of how much the global phishing cost and we compiled them together to make a line graph in Figure 1 [?]. However, they do not specify whether this damage is direct cost

[1] http://www.theregister.co.uk/2004/09/29/phishing_survey/

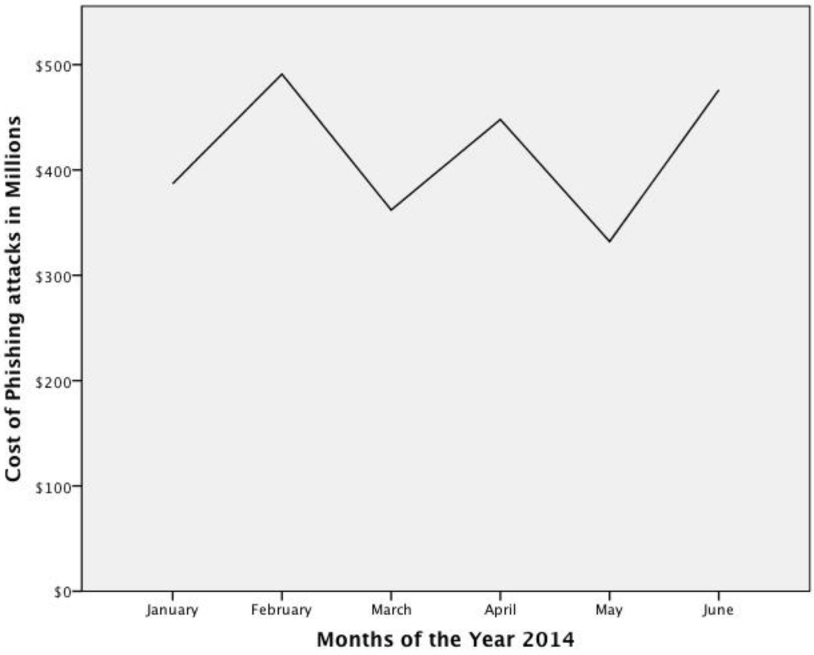


Figure 1: Global phishing cost from January 2014 until June 2014 [?]

or indirect cost. In their study, we can see that there are fluctuation of losses in term of money. Furthermore, this total cost is indeed in accordance with the total cost ranging from $61 million to $3billion per year [? ? ? ?].

1.3 HUMAN FACTOR

Phishing attacks generally aim to manipulate end users to comply phisher's request. Such manipulation in phishing attacks is achieved by social engineering. This means that human element is tightly involved with phishing. But how do phishers compose such deception? How come online users are gullible to these attacks?

Kevin Mitnick, who was obtaining millions of dollars by performing social engineering technique, is plausibly the best known person who had used social engineering technique to carry out his attacks. His book that titled "The art of deception: Controlling the Human Element of Security" [?] has defined social engineering as follows:

> "Using influence and persuasion to deceive people by convincing them that the attacker is someone he is not, or by manipulation. As a result, the social engineer is able to take advantage of people to obtain information, or to persuade them to perform an action item, with or without the use of technology."

> From his definition we can learn that people are the main target of the attack, specifies some of the important tools used by the attackers, such as influence and persuasion.

Cialdini suggests that there are six basic principles of persuasion [?], that is, the technique of making people grant to one's request. These principles include; *reciprocation, consistency, social proof, likeability, authority and scarcity*. Reciprocation is the norm that obligates individuals to repay in kind what they have received, return the favor or adjustment to smaller request [?]. Consistency is a public commitment where people become psychologically become vested in a decision they have made [?][?]. Social proof is when people model the behavior of their peer group, role models, important others or because it is generally "fashionable" [?]. Stajano, et al. suggest people will let their guard down when everybody around them appears to share the same risk [?]. Likeability is when people trust and comply with requests from others who they find attractive or having credibility [? ?]. While it is our human nature not to question authority, it can be used to engender fear, where people obey commands to avoid negative consequences such as losing a privilege or something of value, punishment, humiliation or condemnation [? ?]. Stajano, et al suggest that scarcity is related to time principle, that is, when we are under time pressure to make important choice, we tend to have less reasoning to make decision [?]. We will use these principles as our foundation in synthesizing phishing email corpus with human factor.

Human as the "weakest link" in computer security has been exists and exploited for ages. And yet, security designers blame on users and whine "the system I designed would be secure, if only users were less gullible" [?]. Stajano, et al. stated that "a wise security designer would seek a robust solution which acknowledge the existence of these vulnerabilities as unavoidable consequence of human nature and actively build countermeasures that prevent this exploitation" [?]. With this in mind, the exploration of persuasion principles is congruent with our research goal. Cialdini's six persuasion principles will be the foundation in our research.

1.4 MODUS OPERANDI

As we mentioned earlier, Phishing attack is a subset of identity theft. The modus operandi usually carried out firstly by creating a fake website that spoofs legitimate website such as financial website, either identical or not identical as long as the phishers get responds from unsuspected victims. After that, the phishers will try to trick the potential victim to submit important information such as usernames, passwords, PINs, etc. through a fake website that they have created or through email reply from victims. With the information obtained, they will try to steal money from their victims if target institution is a bank. Phishers employ variety of techniques to trick potential victims to access their fraudulent website. One of the typical ways is by sending illicit email in a large scale claiming to be from legitimate institution. In the email content, they usually imitate an official-looking logo, using good business language style and often also forge the email headers to make it look like originating from legitimate institution. For example, the content of the email is to inform the user that the bank is changing its IT infrastructure, and request urgently that the customer should update their data with the consequence of loosing their money if the action does not take place. When the user click the link that was on the email message, they will be redirected to a fraudulent website, which will prompt the victim to fill in the details of their information. While there are various techniques of phishing attack, we will address the common phases of phishing that we analyzed by literature survey by several studies and also we will address our own phishing phases. These phases are compiled in Table 1.

Table 1: Compilation of phishing phases

J. Hong [?]
1. Potential victims receive a phish 2. The victim may take a suggested action in the message 3. The phisher monetizes the stolen information
Frauenstein, et al. [?]
1. Planning 2. Email Design 3. Fabricated story 4. Threatening tone/Consequences 5. Spoofed website
Wetzel [?]
1. Planning 2. Setup 3. Attack 4. Collection 5. Fraud 6. Post-attack
Tally, et al. [?]
1. The attacker obtains E-mail addresses for the intended victims 2. The attacker generates an E-mail that appears legitimate 3. The attacker sends the E-mail to the intended victims in a way that appears legitimate and obscures the true source 4. The recipient opens a malicious attachment, completes a form, or visits a web site 5. Harvest and exploitation
Emigh [?]

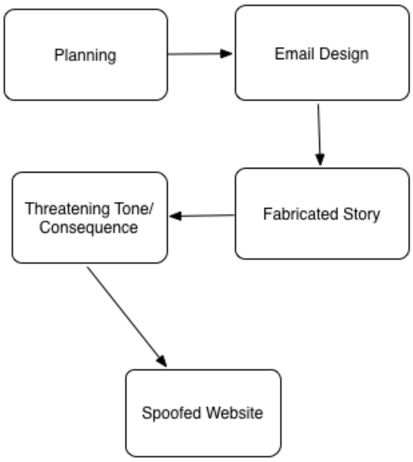


Figure 2: Phishing processes based on Frauenstein[?]

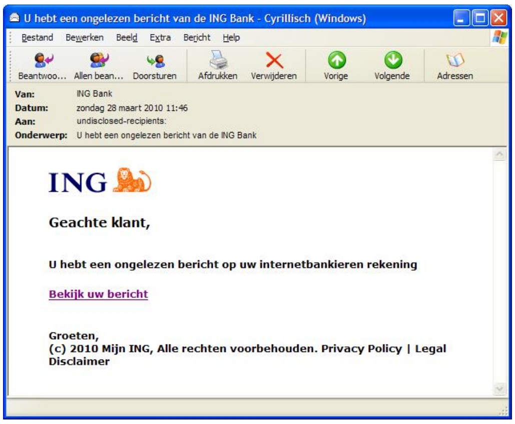
1. A malicious payload arrives through some propagation vector
2. The user takes an action that makes him or her vulnerable to an information compromise
3. The user is prompted for confidential information, either by a remote web site or locally by a Web Trojan
4. The user compromises confidential information
5. The confidential information is transmitted from a phishing server to the phisher
6. The confidential information is used to impersonate the user
7. The phisher engages in fraud using the compromised information

Nero et al. [?]

1. Preparation
2. Delivery of the Lure
3. Taking the Bait
4. Request for Confidential Information
5. Submission of Information
6. Collection of Data
7. Impersonation
8. Financial Gain

Based on the example scenario explained earlier, phishing attacks may consist of several phases. J. Hong [?] argued that there are three major phases. while Frauenstein, et al. [?] suggested that there are five main processes are used to perform phishing attacks based on the perpective of the attacker.

As we illustrated in Figure 2, on the first process is called *Planning*, a phisher usually would do some reconnaissance on how would the attack is executed and what information would be obtained from the victim. On the second process, the phisher would think about the design of the email. This email is desired by the phisher to look as legit as possible to potential victim. For this purpose, target institutions logo, trademark, symbol, etc. are used to make the content look official to the victim. The author called this process as *Email Design*. Figure 3 illustrates the example of fake ING bank logo in a



Van: ING Bank
Datum: zondag 28 maart 2010 11:46
Aan: undisclosed-recipients:
Onderwerp: U hebt een ongelezen bericht van de ING Bank

ING

Geachte klant,

U hebt een ongelezen bericht op uw internetbankieren rekening

Bekijk uw bericht

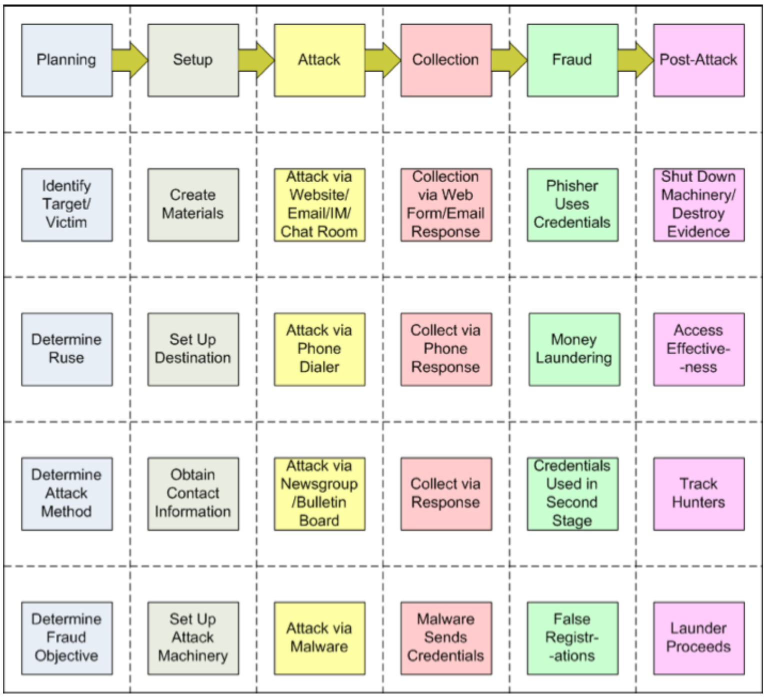
Groeten,
(c) 2010 Mijn ING, Alle rechten voorbehouden. Privacy Policy | Legal Disclaimer

Figure 3: Example of fake ING logo in phishing email

phishing email to create "legitimate" feel[2]. On the third process, the phisher *fabricates* a story to make potential victim think that email is important. To achieve users attention, phisher might build up a story about system upgrade, account hijacked or security enhancement so that the victim would feel obliged to be informed. Evidently, this technique corresponds with Cialdini [**?**] that suggests there are six principles to persuade people to comply with a request. On the fourth process, a phisher usually include *threatening tone* or explain the urgency and consequences if the potential victim chooses not to take action desired by the phisher (for example; account removal, account blocked, etc.). Consequently, users may fear of their account being deleted. The last process involved with fraudulent website that has been created by the phisher. Users may falsely believe to the message given in the email and may click a **URL!** (URL!) that is embedded in the email. Subsequently, the URL would redirect users to a *spoofed website* which may prompt users' sensitive information. Furthermore, the website might be created to be as similar as possible to the target institution's website, so that potential victim may still believe that it is authentic. We will explain more on Cialdini's six basic tendencies of human behavior in generating positive response to persuasion [**?**] in a later section.

Considering that phishing attack is a process, Wetzel [**?**] suggested a taxonomy to make sense of the complex nature of the problem by mapping out a common attacks lifecycle, and a possible set of activities attackers engage in within each phase. The taxonomy is illustrated in Figure 4. We speculated that Wetzel's taxonomy is not analogous with Frauenstein's main phishing processes [**?**]. The difference is that Frauenstein et al. only focus in the design of the attack while Wetzel has added several phases like *Collection*, *Fraud*

[2] http://www.martijn-onderwater.nl/wp-content/uploads/2010/03/ing-phishing.jpg



Planning	Setup	Attack	Collection	Fraud	Post-Attack
Identify Target/ Victim	Create Materials	Attack via Website/ Email/IM/ Chat Room	Collection via Web Form/Email Response	Phisher Uses Credentials	Shut Down Machinery/ Destroy Evidence
Determine Ruse	Set Up Destination	Attack via Phone Dialer	Collect via Phone Response	Money Laundering	Access Effective- -ness
Determine Attack Method	Obtain Contact Information	Attack via Newsgroup /Bulletin Board	Collect via Response	Credentials Used in Second Stage	Track Hunters
Determine Fraud Objective	Set Up Attack Machinery	Attack via Malware	Malware Sends Credentials	False Registr- -ations	Launder Proceeds

Figure 4: Phishing attack taxonomy and lifecycle[**?**]

and *Post-attack*, therefore, Wetzel taxonomy is more holistic in term of phishing.

As we listed Wetzel's taxonomy in Table 1, we explain more of the taxonomy as follows:

1. *Planning*: Preparation carried out by the phisher before continue to the next phase. Example activities include identifying targets and victims, determine the method of the attack, etc.

2. *Setup*: After the target, victim and the method are known, the phisher would craft a platform where the victim's information could be transmitted and stored, for example: fraudulent website/email.

3. *Attack*: Phisher distributes their fraudulent platform so that it can be delivered to the potential victims with fabricated stories.

4. *Collection*: Phisher collects valuable information via response from the victims

5. *Fraud*: Phiser abuses victim's information by impersonates the identity of the victim to the target. For example, A has gained B's personal information to access C so that A can pose as B to access C.

6. *Post-attack*: After the phisher gained profit from the attack and abuse phases, a phisher would not want to be noticed or detected by authority. Thus, phisher might need to destroy evidence of the activities that he/she previously were executed.

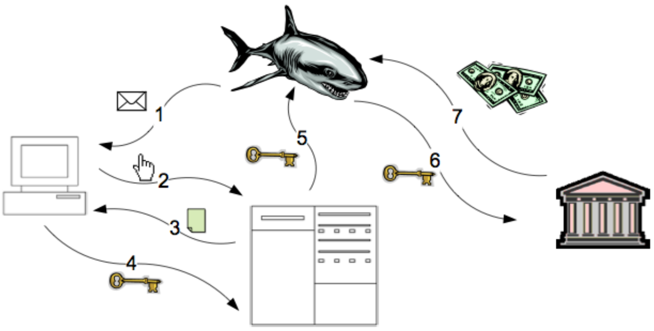


Figure 5: Flow of information in phishing attack [?]

Tally, et al. suggest that there are several phases involved in phishing attack based on the attacker's point of view [?]:

> "*1. The attacker obtains E-mail addresses for the intended victims. These could be guessed or obtained from a variety of sources*" this fits with planning phase.
>
> "*2. The attacker generates an E-mail that appears legitimate and requests the recipient to perform some action*" - fits with design phase.
>
> "*3. The attacker sends the E-mail to the intended victims in a way that appears legitimate and obscures the true source*" - fits with delivery and attack phase.
>
> "*4. Depending on the content of the E-mail, the recipient opens a malicious attachment, completes a form, or visits a web site*" - fits with attack phase.
>
> "*5. The attacker harvests the victim's sensitive information and may exploit it in the future*" - fits with fraud phase.

Aditionally, the phases described by Tally, et al. [?] are comparable with the information flow explained by Emigh[?] represented in Figure 5.

Emigh [?] suggests the information flow of phishing attacks has seven phases. These phases is explained as follow:

> "1. A malicious payload arrives through some propagation vector.
>
> 2. The user takes an action that makes him or her vulnerable to an information compromise.
>
> 3. The user is prompted for confidential information, either by a remote web site or locally by a Web Trojan.
>
> 4. The user compromises confidential information.
>
> 5. The confidential information is transmitted from a phishing server to the phisher.

> 6. The confidential information is used to impersonate the user.
>
> 7. The phisher engages in fraud using the compromised information." [?]

Phishing attack steps performed by the phisher are also addressed by Nero, et al [?]. In their study, a successful phishing attack involves several phases:

> "1. Preparation
>
> 2. Delivery of the Lure
>
> 3. Taking the Bait
>
> 4. Request for Confidential Information
>
> 5. Submission of Information
>
> 6. Collection of Data
>
> 7. Impersonation
>
> 8. Financial Gain" [?]

Based on our analysis by looking at the pattern of other phases from various sources, there is a major similarity between them. Therefore, we would like to define and design our own phase that are integrated with three key components suggested by Jakobsson, et al. [?]. These key components are include *the lure*, *the hook* and *the catch*. As we designed in Figure 6, we synthesized these three components with our phases based on the attacked point of view as follows:

- The lure

1. Phishers prepare the attack
2. Deliver initial payload to potential victim
3. Victim taking the bait

- The hook

4. Prompt for confidential information
5. Disclosed confidential information
6. Collect stolen information

- The catch

7. Impersonates victim
8. Received pay out from the bank

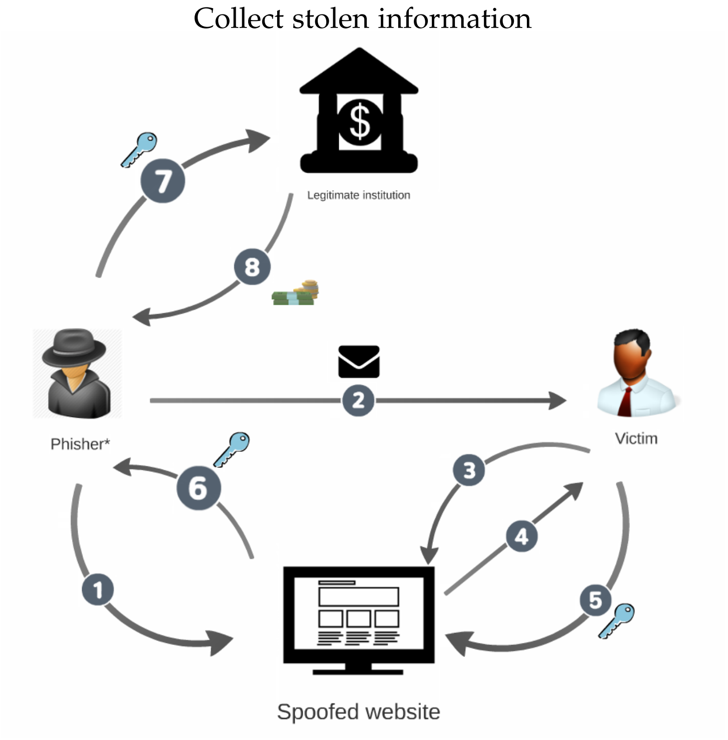


Figure 6: Information flow phishing attack

It is important to know that in the phase 3, there are different scenarios such as; victim might be redirected to a spoofed website, victim may comply to reply the email, victim may comply to open an attachment(s) or victim may comply to call by phone. However, in Figure 6, we have only illustrated the phases if the bait was using a spoofed website as a method.

1.5 TYPES OF PHISHING

In January 2014, 8300 patients data are being compromised in medical company in the US [?]. The data includes names, addresses, date of birth and phone numbers were being stolen. Other than demographic information, clinical information associated with this data was also stolen, including social security numbers. In the April 2014, phishers have successfully stolen US$163,000 from US public school based on Michigan [?]. It has been said that the email prompted to transfer money is coming from the finance director of the school. In March 2014, Symantec has discovered phishing attack aimed at Google drive users [?]. The attack was carried firstly with incoming email asking for opening document hosted at Google docs. Users that have clicked on the link are taken to fraudulent Google login page prompted Google users credentials. Interestingly, the URL seems very convincing because it hosted on Google secure servers. We hypothesized that even more phishing incidents on financial area as well, but sometimes the news is kept hidden due to creditability reason.

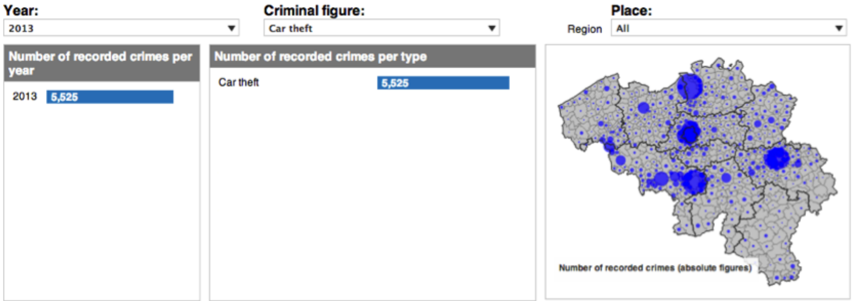
One may ask, what type of phishing are these? What type of phishing commonly used nowadays? Evidently, based on the cost of phishing attacks in Section 1.2, the threat of phishing attacks is still alarming and might be evolving in the future with more sophisticated technique of attacks. For this reason, it might be useful to provide a brief insight on popular variants of phishing that currently exist. We will briefly explain the types of phishing which are the most relevant to our research based on Jakobsson, et al. [**?**]. These types of phishing is strongly related to the phishing defintion that we used, considering phishing is based on the act of deception by the phishers.

1.5.1 *Deceptive phishing*

There are many variations based on deceptive phishing schemes. Typical scenario of deceptive phishing schemes is to send a large amount of illicit emails containing call to action asking recipients to click embedded links [**?**]. These variations include cousin domain attack. For example, legitimate PayPal website addressed as www.paypal.com, this cousin domain attacks confuse potential victims to believe that www.paypal-security.com is a subdivision of the legitimate website due to identical looking addresses. Similarly, homograph attacks create a confusion using similar characters to its addresses. For example, www.paypal.com and www.paypa1.com, both addresses look the same but on the second link, it uses “1” instead of “l”.

Moreover, phishers may embed a login page directly to the email content. This suggests the elimination of the need of end-users to click on a link and phishers do not have to manage an active fraudulent website. IP addresses are often used instead of human readable hostname to redirect potential victim to phishing website and JavaScript is used to take over address bar of a browser to make potential victims believe that they are communicating with the legitimate institution. We will also see few examples of malicious JavaScript on our preliminary analysis section.

Another type of deceptive phishing scheme is rock-phish attacks. They held responsible for half a number of reported incidents worldwide in 2005 [**?**]. These attacks evade email filters by utilizes random text and GIF images which contain the actual message. Rock phish attacks also utilize a toolkit that capable to manage several fraudulent websites in a single domain. Sometimes, deceptive phishing schemes lead to installation of malware when users visit fraudulent website and we will describe malware based phishing scheme in the next section.



Year: 2013

Criminal figure: Car theft

Place: Region All

Number of recorded crimes per year

2013 5,525

Number of recorded crimes per type

Car theft 5,525

Figure 7: Belgium police record on car theft incidents in 2013

1.5.2 *Malware-based phishing*

Generally, malware based phishing refers to any type of phishing which involves installing malicious piece of software onto users' personal computer [?]. Subsequently, this malware is used to gather confidential information from victims instead of spoofing legitimate websites. This type of phishing incorporates malwares such as keyloggers/screenloggers, web Trojans and hosts file poisoning.

1.6 BAD NEIGHBORHOODS ON PHISHING

In the physical world, there are parts of certain area that have higher crime rates than others (eg. Bronx in the US) which called hotspots. Evidently, it is statistically more likely that a crime will occur compared to other locations [?]. To better illustrate this analogy, the police department in Belgium [?] has put up statistical information regarding crime rates in the country. Figure 7 shows an example in 2013; there were up to 5525 car theft incidents recorded and we can see there are certain areas that have higher probability that a car got stolen. For example, Antwerp had 415 incidents whereas Berlare had only 1 car theft incident [?]. The data is not necessarily based on cities, it can be based on the residential area within a city. This holds true that much higher crime rates in a concentrated location compared to any other locations. It is called bad neighborhood.

To reduce the crime rates in a bad neighborhood, it makes sense that the authority should put more enforcement in this location. Moreover, the citizen should avoid this location as much as they can if they want to feel much safer. Evidently, Moura, et al. suggest that the existence of bad neighborhood phenomenon also occurs in the Internet world called "Internet bad neighborhoods" [?]. There are certain networks of Internet infrastructure that contain more malicious activities that other networks. For our preliminary analysis, we will adopt formal definition of Internet bad neighborhoods or Internet Badhoods by [?] which states:

> *"Internet bad neighborhood is a set of IP addresses clustered according to an aggregation criterion in which a number of IP addresses perform a certain malicious activity over a specified period of time"*

Several studies have suggested that the source of the Internet Badhoods tend to be concentrated in certain portions of **IP!** (**IP!**) address space [**?** **?** **?**]. Moura, et al. suggest that Internet Badhoods do not always only based on network prefixes level (e.g. /24, /32, etc..) but it can be aggregated into several levels (ISPs, Countries) [**?**]. Moreover, Internet Badhoods may vary depending on which application exploited. While spam is most likely distributed all around the world, however, phishing Badhoods are most likely concentrated in developed countries (e.g. US) [**?**]. This suggests that phishing sites are required to have more reliable hosts in term of availability, while spams are not. We will see on **??** that consists of our preliminary analysis on phishing Badhoods.

1.7 CURRENT COUNTERMEASURES OF PHISHING ATTACKS

There are various types of phishing countermeasures that implemented in different levels. Purkait has conducted an extensive research in reviewing these countermeasures which are available up until 2012 and their effectiveness [**?**]. He suggests that there is a classification of phishing countermeasures in separate groups and according to Purkait [**?**], these groups are listed as follow:

- Stop phishing at the email level
- Security and password management toolbars
- Restriction list
- Visually differentiate the phishing site
- Two facto and multi channel authentication
- Takedown, transaction anomaly detection, log files
- Anti phishing training
- Legal solution

In addition, Parmar, et al. suggests that phishing detection can be classified into two types; user training approach and software classification approach [**?**]. He illustrated a diagram and a table that summarizes phishing detection as countermeasures in a broad view [**?**]. They also argued the advantages and disadavantages of each category [**?**]. However, as our research mainly focuses in synthesizing phishing email with cialdini's six principles of persuasion [**?**], we will briefly discuss phishing countermeasures such as restriction list group (i.e. Phishtank), machine learning approach (web-based phishing), feature selection in a phishing email, and anti phishing training group (i.e PhishGuru).

1.7.1 *Phishing detection*

Phishtank

One of the common approaches to detect phishing attacks is the implementation of restriction list. As the name suggest, it prevents users to visit fraudulent websites. One of the efforts to achieve restriction list, is to derive phishing URLs from Phishtank. Phishtank is a blacklisting company specifically for phishing URLs and it is a free community web based where users can report, verify and track phishing URLs [**?**]. Phishtank stores phishing URLs in its database and is widely available for use by other companies for creating resriction list. Some of the big companies that are using Phishtank's data includes; Yahoo Mail, McAfee, APWG, Web Of Trust, Kaspersky, Opera and Avira. In this section, we will discuss how the current literatures have to do with phish data provided by Phishtank. The first step to achieve the list of relevant literatures regarding phishtank is by keyword search in Scopus online library. By putting "Phishtank" as a keyword search, it results in 12 literatures. The next step, we read the all the abstracts and conclusions of the resulting keyword search and we decided 11 literatures that are relevant to our research. Lastly, Table 2 summarizes the papers selected and its relevancy with Phishtank

Table 2: Summary phishtank studies

Paper title	**First author**	**Country**	**Relevancy with phishtank**
Evaluating the wisdom of crowds in assessing phishing website [**?**]	Tyler Moore	United Kingdom	Examine the structure and outcomes of user participation in Phishtank. The authors find that Phishtank is dominated by the most active users, and that participation follows a power law distribution and this makes it particularly susceptible to manipulation.
Re-evaluating the wisdom of crowds in assessing web Security [**?**]	Pern Hui Chia	Norway	Examine the wisdom of crowds on web of trust that has similarity with Phishtank as a user based system.

Automatic detection of phishing target from phishing webpage [?]	Gang Liu	China	Phishtank database is used to test the phishing target identification accuracy of their method.
A method for the automated detection of phishing websites through both site characteristics and image analysis [?]	Joshua S. White	New york, US.	Phishtank database is used to perform additional validation of their method. They also collect data from twitter using twitter's API to find malicious tweets containing phishing URLs
Intelligent phishing detection and protection scheme for online transaction [?]	P.A. Barraclough	Newcastle, United Kingdom	Phishtank features is used as one of the input of neuro fuzzy technique to detect phishing website. The study suggested 72 features from Phishtank by exploring journal papers and 200 phishing website.
Towards preventing QR code based attacks on android phone using security warning [?]	Huiping Yao	New Mexico, US.	Phishtank API is used for lookup whether the given QR containing phishing URL in the Phishtank database.
A SVM based technique to detect phishing URLs [?]	Huajun Huang	China	Phishtank database is used as validation resulting 99% accuracy by SVM method, plus the top ten brand names in Phishtank archive is used as features in SVM method.
Socio technological phishing prevention [?]	Gaurav Gupta	Australia	Analyze the Phishtank verifiers (individual/organization) to be used as anti phishing model.
An evaluation of lightweight classification methods for identifying malicious URLs [?]	Shaun Egan	Grahamstown, South Africa	Indicating that lightweight classification methods achieves an accuracy of 93% to 96% when trained data from Phishtank.
Phi.sh/$oCiaL: The phishing landscape through short URLs [?]	Sidharth Chhabra	Delhi, India	Phishtank database is used to analyze suspected phish that is done through short URLs.
Discovering phishing target based on semantic link network [?]	Liu Wenyin	Hong Kong	Phishtank database is used as test dataset to verify their proposed method (Semantic Link Network)

From our literature survey, we know that Phishtank is crowd-sourced platform to manage phishing URLs. For that reason Moore, et al.

aims to evaluate the wisdom of crowds platform accomodated by Phishtank [?]. Moore, et al. suggest that the user participation is distributed according to power law. It uses to model data which frequency of an event varies as a power of some attribute of that event [?]. Power law also applies to a system when large is rare and small is common [3]. For example, in the case of individual weatlh in a country, 80% of the all wealth is controlled by 20% of population in a country. It makes sense that in Phishtank's verification system, a single highly active user's action can greatly impact the system's overall accuracy. Table 3 summarizes the comparison performed by [?] between Phishtank and closed proprietary anti-phishing feeds[4]. Moreover, there are some ways to disrupt Phishtank verification system; submitting invalid reports accusing legitimate website, voting legitimate website as phish, and voting illegitimate website as not phish. While all the scenarios described are for the phishers' benefit, the last scenario is more direct and the first two actions rather subtle intention to undermine Phishtank credibility.

To put it briefly, the lesson of crowd sourced anti-phishing technology such as Phishtank is that the distribution of user participation matters. It means that if a few high value participants do something wrong, it can greatly impact overall system [?]. Also, there is a high probability that bad users could also extensively participate in submitting or verifying URLs in Phishtank.

Machine learning approach in detecting spoofed website

The fundamental of phishing detection system would be to distinguish between phishing websites and the legitimate ones. As we previously discussed, the aim of phishing attack is to gather confidential information from potential victims. To do this, phishers often prompt for this information through fraudulent websites and masquerade as legitimate institutions. It does not make sense if phishers created them in a way very distinctive with its target. It may raise suspicions with result of unsuccessful attack. To put it another way, while it might be true, we speculated that most of the phishing websites are mostly identical with its legitimate websites as target to reduce suspiciousness from potential victim.

In contrast of one of blacklisting technique we saw in Phishtank that heavily depend on human verification, researchers make use of machine learning based technique to automatically distinguish between phishing and legitimate either websites or email. Basically, machine-learning system is a platform that can learn from previous data and predict future data with its classification, in this case, phishing and legitimate. In order for this machine to learn from data, there

[3] http://kottke.org/03/02/weblogs-and-power-laws

[4] The author conceals the identity of the closed proprietary company

Phishtank	Proprietary
10924 URLs	13318 URLs
8296 URLs after removing duplication	8730 URLs after removing duplication
Shares 5711 URLs in common 3019 Unique to the company feeds while 2585 only appeared in Phishtank	
586 rock-phish domains	1003 rock phish domains
459 rock phish domains found in Phishtank	544 rock phish domains not found in Phishtank
Saw the submission first	11 minutes later appear on the feed
16 hours later after its submission for verification (voting based)	8 second to verified after it appears
Rock phish appear after 12 hours appeared in the proprietary feed and were not verified for another 12 hours	

Table 3: Comparison summary [?]

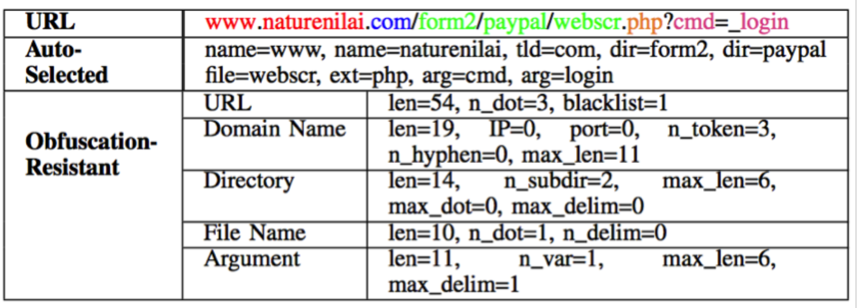
should be some kind of inputs to classify the data, it is called features or characteristics.

Furthermore, there are also several learning algorithms to classify the data, such as, logistic regression, random forest, neural networks and support vector machine. However, for the sake of simplicit of our research, we will not discuss about the learning algorithm that is currently implemented. We will only introduce three features that are used in machine learning based detection.

There are vast amount of features to utilize machine learning to detect phishing attack. Literatures are selected by keyword search such as "phishing + detection + machine learning". We analyze three features: lexical feature, host-based feature and site popularity feature. Each of these features will be introduced briefly as follows.

- Lexical features

Lexical features (URL based features) are based on the analysis of URL structure without any external information. Ma, et al. suggest that the structure URL of phishing may "look" different to experts [?]. These features include how many dots exist, the length, how deep the path traversal do the URL has or if there any sensitive words present in a URL. For example the URLs https://www.paypal.com and http://www.paypal.com.example.com/ or http://login.example.com/ www.paypal.com/, we can see that the domain paypal.com positioned



URL	www.naturenilai.com/form2/paypal/webscr.php?cmd=_login	
Auto-Selected	name=www, name=naturenilai, tld=com, dir=form2, dir=paypal file=webscr, ext=php, arg=cmd, arg=login	
Obfuscation-Resistant	URL	len=54, n_dot=3, blacklist=1
	Domain Name	len=19, IP=0, port=0, n_token=3, n_hyphen=0, max_len=11
	Directory	len=14, n_subdir=2, max_len=6, max_dot=0, max_delim=0
	File Name	len=10, n_dot=1, n_delim=0
	Argument	len=11, n_var=1, max_len=6, max_delim=1

Figure 8: Example lexical features [?]

differently, with the first one being the benign URL. Figure 8 shows an example analysis of lexical features in a phishing URL [?].

Lexical features analysis may have performance advantage and reduces overhead in term of processing and latency, since it only tells the machine to learn URL structure. 90% accuracy is achieved when utilizing lexical features combined with external features such as WHOIS data [?]. Egan, et al. conducted an evaluation of lightweight classification that includes lexical features and host based features in its model [?]. The study found that the classification based on these features resulted in extremely high accuracy and low overhead. Table 4 lists the existing lexical features that are currently implemented by two different studies [? ?]. However, Xiang, et al.[?] pointed out that URLs structure could be manipulated with little cost, causing the features to fail. For example, attackers could simply remove embedded domain and sensitive words to make their phishing URLs look legitimate. Embedded domain feature examines whether a domain or a hostname is present in the path segment [?], for example, http://www.example.net/pathto/www.paypal.com. Suspicious URL feature examine whether the URL has "@" or "-", the present of "@" is examined in a URL because when the symbol "@" is used, the string to the left will be discarded. Furthermore, according to [?], not many legitimate websites use "-" in their URLs. There are also plenty of legitimate domains presented only with IP address and contains more dots. Nevertheless, lexical analysis would be suitable features to use for first phase analysis in a large data [?].

- Host based features

Since phishers often hosted phishing websites in less reputable hosting services and registrars, host-based features are needed to observe on the external sources (WHOIS information, domain information, etc.). A study suggests host-based features have the ability to describe where phishing websites are hosted, who owns them and how they are managed [?]. Table 5 shows the host-based features from three studies that are currently used in machine learning phishing detection. These studies are selected only for example comparison.

Haotian Liu, et al. [?]	**Guang Xiang, et al. [?]**
- Length of hostname Length of entire URL - Number of dots - Top-level domain - Domain token count - Path token count - Average domain token length of all dataset - Average path token length of dataset - Longest domain token length of dataset - Longest path token length of dataset - Brand name presence - IP address presence - Security sensitive word presence	- Embedded domain - IP address presence - Number of dots - Suspicious URL - Number of sensitive words - Out of position top level domain (TLD)

Table 4: Existing lexical features [? ?]

Justin Ma, et al.[? ?]	**Haotian Liu, et al. [46][?]**	**Guang Xiang, et al. [?]**
- WHOIS data - IP address information - Connection speed - Domain name properties	- Autonomous system number - IP country - Number of registration information - Number of resolved IPs - Domain contains valid PTR record - Redirect to new site - All IPs are consistent	- Age of Domain

Table 5: Host-based features [? ? ? ?]

Guang Xiang, et al. [?]	**Haotian Liu, et al. [?]**
- Page in top search results - PageRank - Page in top results when searching copyright company name and domain - Page in top results when searching copyright company name and hostname	- Number of external links - Real traffic rank - Domain in reputable sites list

Table 6: Site popularity features [? ?]

Each of these features does matter for phishing detection. However, as our main objective is synthesizing cialdini's principle with phishing emails, we will not describe each of these features in detail. It is noteworthy that some of the features are subset of another feature, for instance, autonomous system number (ASN), IP country and number of registration information are derived from WHOIS information. Nevertheless, we will only explain few of them that we assume the most crucial.

1. WHOIS information: Since phishing websites and hacked domains are often created at relatively young age, this information could provide the registration date, update date and expiration date. Domain ownership would also be included; therefore, a set of malicious websites with the same individual could be identified.

2. IP address information: Justin Ma, et al. used this information for identify whether or not an IP address is in blacklist [? ?]. Besides the corresponding IP address, it provides records like nameservers and mail exchange servers. This allows the classifier to be able to flag other IP addresses within the same IP prefix and ASN.

3. Domain name properties: these include time to live (TTL) of DNS associated with a hostname. PTR record (reverse DNS lookup) of a domain could also be derived whether it is valid or not.

- Site popularity features

Site popularity could be an indicator whether a website is phishy or not. It makes sense if a phishing website has much less traffic or popularity than a legitimate website. According to [?], some of the features indicated in Table 6 are well performed when incorporated with machine learning system.

1. Page in top search results: this feature originally used by [?] to find whether or not a website shows up on the top N search result. If it is not the case, the website could be flagged as phishy since phishing websites have less chance of being crawled [?]. We believe this feature is similar to Number of external links feature since both of them are implying the same technique.

2. PageRank: this technique is originally introduced by Google to map which websites are popular and which are not, based on the value from 0 to 10. According to [?], the intuitive rationale of this feature is that phishing websites are often have very low PageRank due to their ephemeral nature and very low incoming links that are redirected to them. This feature similar to Real traffic rank feature employed by [?] where such feature can be acquired from alexa.com.

3. Page in top results when searching copyright company name and domain/hostname features are complement features of Page in top search results feature with just different queries. Moreover, we believe they are also similar to Domain in reputable sites list feature since they are determining the reputation of a website. The first two features can be identified by querying google.com [?] and the latter feature can be obtained from amazon.com [?].

Stop phishing at email level

In order to stop phishing at email level, phishing email properties or features should be investigated. Chandrasekaran, et al. and Drake, et al [? ?] specify the structure of phishing emails properties as follows:

1. Spoofing of online banks and retailers. Impersonation of legitimate institutions may created in the email level. Phishers may design a fake email to resemble the reputable company to gain users trust.

2. Link in the text is different from the destination. A link(s) contained in the email message usually appear to different than the actual link destination. This method used to trick users to believe that the email is legitimate.

3. Using IP addresses instead of URLs. Sometimes phishers may hide the link in the message by presenting it as IP address instead of URL.

4. Generalization in addressing receipents. As phishing emails are distributed by large number of recipients, the email often is not personalized, unlike the legitimate email that address its recipient by personalized information such as the last four digits of account information.

5. Usage of well-defined situational contexts to lure victims. Situational contexts such as false urgency and threat are a common method to influence the decision making of the recipients.

Moreover, Ma, et al. experimented with seven properties to consider in a phishing emails consist of the toal number of links, total numbers of invisible links, whether the link that appears in the message is different than the actual destination, the existence of forms, whether scripts exist within an email, total appearance of blacklisted words in the body and the total appearance of blacklisted words in the subject [**?**]. Based on this survey, we established phishing email properties as variables in order to classify our data in **??**.

1.7.2 *Phishing prevention*

Phishing attacks aim to by-pass technological countermeasures by manipulating users' trust and can lead to monetary losses. Therefore, human factors take a big part on the phishing taxonomy, especially in the organizational environment. Human factor in phishing taxonomy comprised of education, training and awareness [**?**]. Figure 9 illustrates where human factor takes part on phishing threats [**?**]. User's awareness of phishing has been explored by several studies [**? ? ? ? ? ?**] as preventive measure against phishing attack. According to ISO/IEC 27002 [**?**][**?**], it has been shown that information security awareness is important and it has been critical success factors to mitigate security vulnerabilities that attack user's trust. One approach to hopefully prevent phishing attack was by implementing anti phishing warning/indicator. Dhamija, et al suggest that users often ignore security indicators thus makes them ineffective [**?**]. Even if users notice the security indicators, they often do not understand what they represent.

Moreover, the inconsistency of positioning on different browsers makes them much difficult to identify phishing [**?**]. Evidently, Schechter, et al. pointed out that 53% of their study participants were still attempting to provide their confidential information, even after their task was interrupted by strong security warning [**?**]. Therefore, these suggest that an effective phishing education must be added as a complementary strategy to complete technical anti-phishing measure as a strong remedy to detect phishing websites or emails.

Phishing education for online users often by instructing not to click links in an email, ensure that SSL is present and to verify that the domain name is correct before giving information, and other similar education. This traditional practice evidently has not always effective [**?**]. One may ask what makes phishing education effective? A study suggests that in order online users to be aware of phishing threats, is to really engage them to so that they understand how vulnerable they are [**?**]. To do this, simulated phishing attacks often performed

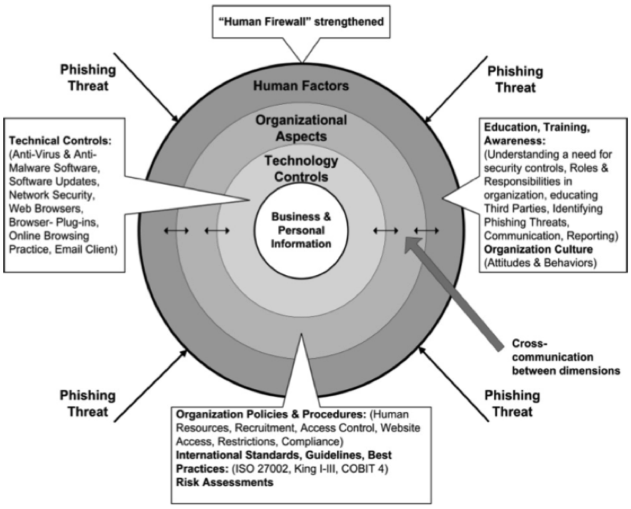
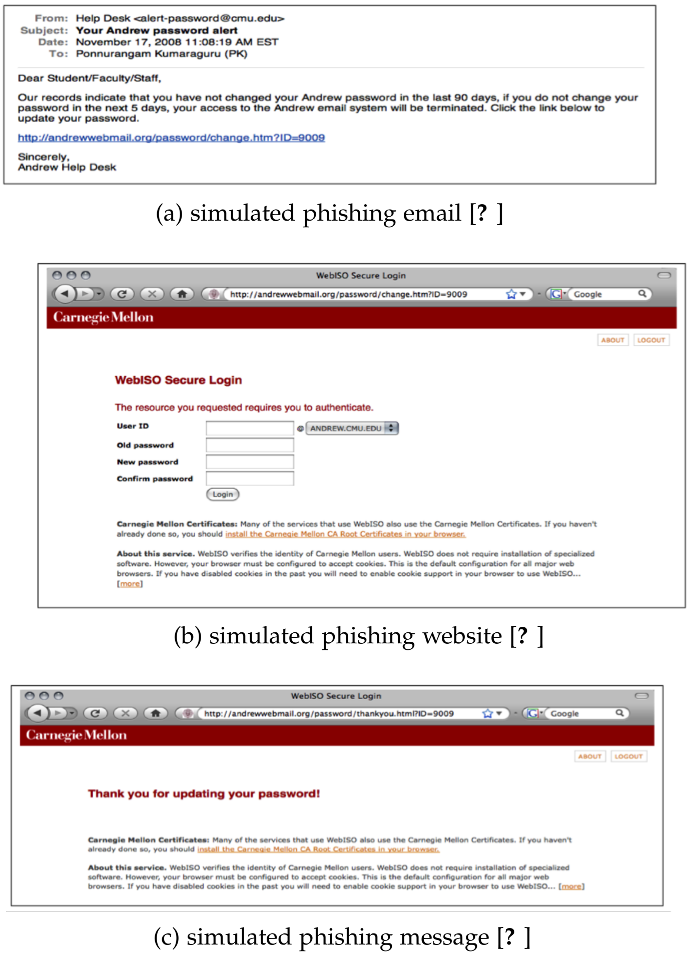

"Human Firewall" strengthened
Phishing Threat
Phishing Threat
Human Factors
Organizational Aspects
Technology Controls
Business & Personal Information
Technical Controls: (Anti-Virus & Anti-Malware Software, Software Updates, Network Security, Web Browsers, Browser- Plug-ins, Online Browsing Practice, Email Client)
Education, Training, Awareness: (Understanding a need for security controls, Roles & Responsibilities in organization, educating Third Parties, Identifying Phishing Threats, Communication, Reporting) Organization Culture (Attitudes & Behaviors)
Cross-communication between dimensions
Phishing Threat
Phishing Threat
Organization Policies & Procedures: (Human Resources, Recruitment, Access Control, Website Access, Restrictions, Compliance) International Standards, Guidelines, Best Practices: (ISO 27002, King I-III, COBIT 4) Risk Assessments

Figure 9: Holistic anti-phishing framework [?]

internally in an organization. Figure 10 shows a simulated phishing email and website carried out by Kumaraguru, et al. from PhishGuru [?]. As a result, this scenario puts them in the ultimate teachable moment if they fall for these attacks.

Phishguru is a security training system operated by Wombat security technology that teaches users not to be decieved by phishing attempts by simulation of phishing attacks[?]. They claimed Phishguru provides more effective training than traditional training as it is designed to be more engaging. Figure 11 illustrates how embedded phishing training was presented by PhishGuru.

Kumaraguru, et al. investigates the effectiveness of embedded training methodology in a real world situation [?]. Evidently, they indicated that even after 28 days after training, users trained by PhishGuru were less likely to click the link presented in the simulated phishing email than those who were not trained. They also find that users who trained twice were less likely to give information to simulated fraudulent website than users who were trained once. Moreover, they argue that the training does not decrease the users' willingness to click on the links from legitimate emails; it means that less likely a trained user did a false positive when he or she requested to give information from true legitimate emails [?]. This suggests that user training strategy as an effective phishing education in order to improve phishing awareness especially in organizational environment.



From: Help Desk <alert-password@cmu.edu>
Subject: **Your Andrew password alert**
Date: November 17, 2008 11:08:19 AM EST
To: Ponnurangam Kumaraguru (PK)

Dear Student/Faculty/Staff,

Our records indicate that you have not changed your Andrew password in the last 90 days, if you do not change your password in the next 5 days, your access to the Andrew email system will be terminated. Click the link below to update your password.

http://andrewwebmail.org/password/change.htm?ID=9009

Sincerely,
Andrew Help Desk

(a) simulated phishing email [?]

WebISO Secure Login

The resource you requested requires you to authenticate.

User ID @ ANDREW.CMU.EDU
Old password
New password
Confirm password
Login

Carnegie Mellon Certificates: Many of the services that use WebISO also use the Carnegie Mellon Certificates. If you haven't already done so, you should install the Carnegie Mellon CA Root Certificates in your browser.

About this service. WebISO verifies the identity of Carnegie Mellon users. WebISO does not require installation of specialized software. However, your browser must be configured to accept cookies. This is the default configuration for all major web browsers. If you have disabled cookies in the past you will need to enable cookie support in your browser to use WebISO... [more]

(b) simulated phishing website [?]

Thank you for updating your password!

Carnegie Mellon Certificates: Many of the services that use WebISO also use the Carnegie Mellon Certificates. If you haven't already done so, you should install the Carnegie Mellon CA Root Certificates in your browser.

About this service. WebISO verifies the identity of Carnegie Mellon users. WebISO does not require installation of specialized software. However, your browser must be configured to accept cookies. This is the default configuration for all major web browsers. If you have disabled cookies in the past you will need to enable cookie support in your browser to use WebISO... [more]

(c) simulated phishing message [?]

Figure 10: Simulated phishing attack [?]

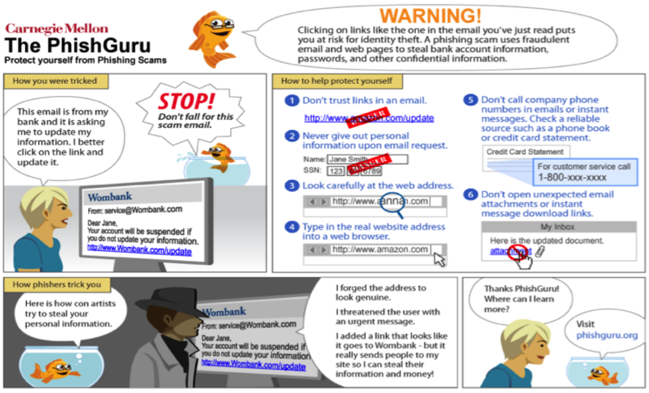

Carnegie Mellon
The PhishGuru
Protect yourself from Phishing Scams
WARNING!
Clicking on links like the one in the email you've just read puts you at risk for identity theft. A phishing scam uses fraudulent email and web pages to steal bank account information, passwords, and other confidential information.
How you were tricked
This email is from my bank and it is asking me to update my information. I better click on the link and update it.
STOP! Don't fall for this scam email.
Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.Wombank.com/update
How to help protect yourself
1 Don't trust links in an email.
http://www.amazon.com/update
DANGER
2 Never give out personal information upon email request.
Name: Jane Smith
SSN: 123 45 6789
DANGER
3 Look carefully at the web address.
http://www.aannan.com
4 Type in the real website address into a web browser.
http://www.amazon.com
5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.
Credit Card Statement
For customer service call 1-800-xxx-xxxx
6 Don't open unexpected email attachments or instant message download links.
My Inbox
Here is the updated document.
attachment
How phishers trick you
Here is how con artists try to steal your personal information.
Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.Wombank.com/update
I forged the address to look genuine.
I threatened the user with an urgent message.
I added a link that looks like it goes to Wombank - but it really sends people to my site so I can steal their information and money!
Thanks PhishGuru! Where can I learn more?
Visit phishguru.org

Figure 11: Embedded phishing training [?]