Phishing as a Service: Designing an ethical way of mimicking targeted phishing attacks to train employees

K.C. Meijdam
TU Delft
Jaffalaan 5
2628 BX Delft
0031 6 00000000
Katrien.meijdam@gmail.com

Dr. ir. W. Pieters
TU Delft
Jaffalaan 5
2628 BX Delft
0031 15 27 89898
W.Pieters@tudelft.nl

Prof. dr. ir. J. van den Berg
TU Delft
Jaffalaan 5
2628 BX Delft
0031 15 27 82794
J.vandenBerg@tudelft.nl

ABSTRACT
Organisations want to protect themselves against targeted phishing attacks, and mimicking those in order to train their employees in an important measure to do so. However, previous experiments have shown that this is ethically sensitive and might lead to feelings of deception and harm at the side of the employees. In this paper, the design of an ethical way to implement mimicked targeted phishing attacks to train employees is described. The design process is based on Value Sensitive Design [1] and Values at Play [2], and use the notion of Value Hierarchies [3] to make the design process explicit and discussable.

Categories and Subject Descriptors
D.4.6 [Operating systems]: Security and Protection

General Terms

Keywords
Value Sensitive Design, targeted phishing, mimicked phishing training

1. INTRODUCTION
A phishing email appears to be a benign email of a trustful sender, tricking the receiver into sharing personal information. Cisco [4] conducted research with organizations worldwide across a broad range of industries. The study revealed that untargeted phishing is decreasing: between 2010 and 2011, annual returns for untargeted email-based attacks fell from $1.1 billion to $500 million. Meanwhile, the number of sent spam email fell from 300 billion messages per day to 40 billion. Instead of sending hundreds of thousands phishing emails to random receivers, senders now often carefully target specific individuals, groups, or employees of a company or organization [5].

Such a targeted attack, incorporating context information, will be referred to as targeted phishing. Personal information on targets is often found out in the open, on social networks like Facebook and Twitter. Brown et al. found that targeted emails could be send to approximately 85% of users [6]. Symantec [7] found a 42% increase in the targeted-attack rate in 2012 when compared to the previous year. Targeted phishing is also being used against high-level targets, like managers and executives [5], [8]. In business context high level targets can be high-level employees or IT professionals with high access rights.

Phishing is usually considered to be a problem in the private domain, but an increasing number of organizations understand that they are also affected when their employees are targeted. Giving the rise of targeted attacks, in order to protect their organizations from intrusions using targeted phishing, organizations should include targeted phishing as part of the training for their employees.

Much research is done on the education of users to get them to behave more securely [9]–[12]. Embedded training is believed to be the most effective training, as opposed to sending notices [13]: Employees fallen for a mimicked attack immediately receive educational material helping them to avoid doing so next time.

Deloitte offers this service, called ‘Phishing as a Service’, abbreviated as PhaaS. In coordination with a decision maker at the client, mimicked attacks, combined with embedded training, will be send to the client’s employees. This is often repeated over a period of time, with email messages increasing in difficulty to recognize as a phishing attempt. The awareness of employees is raised in the process and it provides metrics that allow for comparison with organizations alike [14]. The metrics include the portion of employees opening the email, clicking the link, and providing credentials.

In order to train employees to recognize targeted phishing attacks, these kinds of attacks will also have to be mimicked. Previous studies using personal information in mimicked phishing emails [15] faced negative reactions of participants, who were not informed upfront about their partaking in the experiment.

2. METHODOLOGY
As the previously mentioned experiment by Jagatic et al. [15] illustrates, a challenge while experimenting with mimicking phishing, especially those involving context information, is that it might provoke feelings of displeasure. Reactions of participants of the experiment entailed feelings of anger, denial, or participants...
who didn’t understand the methodology of the researchers and accused them of hacking. These reactions are understandable when the setup of the experiment is studied: participants were not informed upfront, or asked for consent. This kind of setup will not contribute to the goal of mimicking phishing in a business environment: creating awareness and teaching employees signs on which they can recognize phishing. That, and the fact that the relationship between employer and employee needs to be preserved, contribute to the need for a mimicked phishing attack that deals with those ethical issues.

### 2.1 Value Sensitive Design and Values at Play

For a successful design of a procedure to introduce mimicked targeted phishing, it is important to respect the values of all involved stakeholders, such as the targets, the managers and the executers of the attacks. So, it is evident that values are important to shape the design, but how to proceed is unclear. The relationship between values and technology is often described, yet only a few approaches aim to provide methodological guidelines for systematically identifying and accounting for values in the development of systems [16]. Friedman et al. [17] provide an overarching framework that supports the involvement of values dimensions in design work, called Value Sensitive Design (VSD). This is combined with the framework of Values at Play (VAP) [2], which is based on the same investigations but provide more guidance in the activities to be undertaken.

Both frameworks describe three different kinds of investigations that should help with identifying the values that are relevant and how they influence the design process. They are conceptual, empirical, and technical investigations. The essence of the investigations is the integrative and iterative way of combining insights from these three ways.

To structure the design process, the activities mentioned in the VAP framework are paired with the development of a Value Hierarchy, a hierarchy structure of values, norms and design requirements introduces by Van de Poel [3].

### 2.2 Value hierarchies

Once the relevant values are identified, they need to be used in the design process. The VAP framework provides three activities that should be carried out in an iterative way: discovery of values, translation of values and verifying the design [18]. Putting the values in concrete terms is very important, but difficult, because values only emerge indirectly, as shown by the reactions of the Jagatic experiment. To make sure the values are taken into account in the final design, they should be translated in design requirements. This particular step is a rarely discussed aspects of VSD, while it is very important to the shaping of the design process and the success of the outcome [3].

In his paper ‘Translating values into design requirements’ [3], Van de Poel introduces the notion of values hierarchy, as depicted in Figure 1. He stated that values have a hierarchal nature, and design requirement constitute the most concrete layer. The upper layer of values is translated into requirements by means of an intermediate layer, where the norms of a user are placed. The norms are sometimes referred to as objectives, goals and constraints.

For the design process of MTPA, both methodologies are combined; Value Hierarchies are constructed, and the focus on verification by Flanagan et al. is endorsed. The steps that are used during the design of a step-by-step guide to implement MPTA in an ethical way are presented in Table 1.

#### Table 1: Steps of the design process of MPTA

<table>
<thead>
<tr>
<th>Step #</th>
<th>Description of the step</th>
<th>Investigation type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Map the current PhaaS practice and the services of the competitors</td>
<td>Technical</td>
</tr>
<tr>
<td>Step 2</td>
<td>Extracting norms from case studies</td>
<td>Empirical</td>
</tr>
<tr>
<td>Step 3</td>
<td>Linking norms to values</td>
<td>Conceptual</td>
</tr>
<tr>
<td>Step 4</td>
<td>Set up requirements and solve value conflicts</td>
<td>Technical</td>
</tr>
<tr>
<td>Step 5</td>
<td>Make provisional design</td>
<td></td>
</tr>
<tr>
<td>Step 6</td>
<td>Verification and validation</td>
<td>Empirical</td>
</tr>
</tbody>
</table>

### 3. ANALYSIS

#### 3.1 Technical investigations

As explained by Friedman et al. (2013), the Value Sensitive Design approach is based on the belief that technology is not value-neutral. The functionalities provided by technology support or hinder values. The technical investigations try to identify the room for decision making provided by technology, by studying the services of many companies offering mimicked phishing attacks. It has also taken a closer look to the current Phishing as a Service proposition offered by Deloitte.

By using the experience of Deloitte professionals, the decisions that have to be made by the client are identified. They are presented in an ‘empty framework’, as depicted in Figure 2. This framework will function as the starting point for the design process.

The aim of this paper is to provide guidance for anyone deliberating those decisions. This will be by taking into account values that are important to the involved stakeholders. In this chapter, the following was found:

- Some competitors focus on singling out risky employees in order to address them personally, which is violating their privacy.
In order to come to the norms, all normative statements that are found in the interview transcripts, are extracted and grouped. For each statement, a more general norm was formulated. When possible, a norm that was already mentioned was applied, as to group the norms. This led to 7 norms:

- Employees should not feel deceived
- It should not damage the self-image of employees too much
- It should not be possible to trace back results to individual employees
- It should be a positive experience for employees
- It should create more support for awareness exercises
- It should create (lasting) security awareness
- It should be open and transparent
- It should have a high response rate

3.3 Conceptual investigations

During the conceptual investigations, the norms are linked to overarching values. The values are then conceptualized, and perhaps additional norms will be the result. This process is an iterative one, so sometimes a norm is linked to a value of the other way around, or is revised when additional insights are gained. This is the first step in creating the value hierarchies; the results of these investigations are the upper two levels of the value hierarchies, depicted in Figure 3, page 2.

When merging the separate value hierarchies, an overarching insight emerges. In the case of three values, an overarching value is added. Both ‘top management support’, ‘self-efficacy’, and ‘reach many employees’ are related to security awareness. When considering the other values, there seems to emerge another overarching value: the wellbeing of the employee. ‘Privacy’, ‘good employment practice’, and both ‘minimizing harm’ and ‘autonomy’ are concerned with the feelings of the employee. All values can be associated with one of the two overarching values. This is also convenient to identify the dilemmas. They emerge when trying to translate the values into requirements.

4. DILEMMAS IN VALUES

The resulting upper layers of the value hierarchy are translated into design requirements, constructing the lowest level of the Value Hierarchy. Converting the norms into requirements is called specification [3]. As suspected, the translation of norms into requirement has led to four contradicting requirements:

1. ‘Do not warn employees upfront about the exercise’ vs ‘Inform employees upfront about nature of the exercise, expected day of delivery and procedure of the exercise’
2. ‘Do not offer an opt-out for employees’ vs ‘Ask employees to indicate whether they want to participate’
3. ‘Display direct feedback after an employee falls for a mimicked phishing attack’ vs ‘Display no direct debrief message at end of attack’
4. ‘Include many indicators to recognize MTPA’ vs ‘Base scenario on real incoming phishing attacks’ vs ‘Difficulty of attack should match users’ ability’

The contradicting requirements are all investigated by looking at which norms and values they stem from. It can be concluded that

---

**Figure 2: Empty framework mimicked phishing**

- Generally, there is little attention to employee wellbeing among competitors.
- By providing a portal with direct access to clients, some competitors give freedom to the client, and lose control about the way the phishing is executed.
- Deloitte offers other security test as well, as opposed to competitors who focus solely on phishing.
- Portal of Deloitte allows for the preservation of privacy of the targets, because it is not possible to trace back results to individuals.
- Deloitte receives no credentials of employees on their servers. This results in the fact that it is not possible to check whether the credentials are correct.

**3.2 Empirical investigations**

Norms are statements about how things should be, and are more specific then values, and less then design requirements. Norms say something about the actions that should or should not be taken to pursue a value [3].

In order to get a clear and complete picture of the possible choices one can make, four case studies are investigated and interviews are held with the involved decision makers. The cases describe the implementation of mimicked phishing in different companies. All have been executed at clients of Deloitte, using the Phishing as a Service proposition. The interviews focus on the decisions made and on the reasons for those decisions.
all conflicting requirements can be traced back to the tension between the two overarching values ‘security awareness’ and ‘well-being of the employee’. Apparently, striving for one causes conflict in the other one.

If one would strive only for security awareness, the exercise would look the following way: employees would not be aware of the exercise, would consequently all participate and the first time he or she are made aware of the exercise is right after he or she has fallen for it. The difficulty of the exercise can either be very easy, to strengthen his feelings of empowerment, or be very realistic, to provide a truthful view of the threat of phishing to the organization.

If prioritizing the well-being of employees over other values, the exercise would look like this: All employees will be extensively informed about the exercise, would have to opt-in to participate, would not notice when they would have fallen for the attack, because a benign message is displayed after they walked through all the steps. And finally, the difficulty of the attack matches the ability of the employee, so that optimal learning is achieved.

Resolving the value conflicts is a difficult thing to do, and literature provides little grip on how to resolve them. Therefore, a careful consideration is made, in accordance with the experts of Deloitte.

First, the two requirements related to the consent of the employees are considered. It seems that they are at the opposite ends of the spectrum and a golden mean might be found. Both bring along advantages, but both also have severe disadvantages. To not offer an opt-out, forces the few employees who might be against this exercise, into something they do not agree with, causing a lot of trouble. And to ask everybody to opt-in, will probably lead to a low share of employees participating, because employees might be too lazy to go through the trouble of opting in. The golden mean would therefore be to offer an opt-out to all employees. This would filter out employees who really got a problem with the exercise, while keeping indifferent employees on board.

The decision to include an opt-out forces another dilemma; the one regarding the information upfront. In order to offer employees an opt-out, they should be told about the exercise. This would also be the right thing to do with regard to being a good employer. Informing employees upfront will have the negative side-effect that the exercise will not mimic a real phishing attack realistically.

Because the employees are aware of the exercise happening, it makes sense to inform them when they have fallen for the attack with a direct debriefing message. In the one case study where this was not done, the employees were also not informed upfront, so the decision maker feared a too great shock effect when employees were suddenly confronted with the exercise. Because the employees are informed upfront, direct debriefing will be implemented as well. This also provides the opportunity to ask employees not to tell their colleagues, making the exercise more realistic.

Finally, the difficulty of the exercise should be discussed. The best option in this case, is to make the exercise matching with the ability of the employees. Because the exercise can and should be repeated periodically, the difficulty can be increased up to a level that matches real incoming attacks. This will balance the number of times employees experience failure and success. When they fail to recognize the attack, they will be provided with immediate feedback. When they recognize it and report it, their confidence rises and their feeling of control is strengthened. If employees are informed about the adjusted level of difficulty, they will see falling for the attack not per se as failure, because they know the level is adjusted to their ability, so occasional failure is part of the deal. Furthermore, they cannot compare their performance to other employees, because they do not know the level of difficulty they encountered. This option combines the pros and cons of the other requirements, because employees are empowered and eventually, a realistic view of the threat can be given.

The solutions to the value conflicts are presented in Table 2.

Table 2: Proposed solution to value conflicts

<table>
<thead>
<tr>
<th>Focus on security awareness</th>
<th>Focus on well-being of employee</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>No information about exercise to employees upfront</td>
<td>Inform employees upfront about nature of the exercise, expected day of delivery and procedure of the exercise</td>
<td>Inform upfront</td>
</tr>
<tr>
<td>No opt-out for employees</td>
<td>Opt-in for employees</td>
<td>Opt-out for employees</td>
</tr>
<tr>
<td>Direct feedback after an employee falls for a mimicked phishing attack</td>
<td>No direct debriefing message at end of attack</td>
<td>Direct debriefing</td>
</tr>
<tr>
<td>Two options: include many indicators to recognize MTTPA or base scenario on real incoming phishing attacks</td>
<td>The difficulty of the attack should match users’ ability</td>
<td>Difficulty of attack matches users’ ability</td>
</tr>
</tbody>
</table>

5. DESIGN

Based on the requirement, the design is made. It is depicted in Error! Reference source not found.. One of the most important changes is the introduction of a separate phase, called ‘monitoring’, that is important during the whole implementation of mimicked phishing. It is introduced because of the need of a formal procedure to report phishing attacks by employees, to provide positive feedback and to get a clearer picture of the current phishing attacks that are threatening the organization.

The step-by-step guide differs from the original framework in the amount of guidance is given the clients: where first many questions needed answering, now these are answered. The answers are based on an extensive design process, where values have been taken into account. The biggest alteration is the advice to set up a procedure for employees to report alleged phishing attacks the encounter. This procedure will give the employees a feeling of empowerment, and at the same time enable to possibility to provide direct feedback when succeeding in recognizing the MTTPA. Formerly, only direct feedback could be given when an employee had fallen for the attack, making the exercise a negative experience.

Another big alteration compared to the current process, is the fact that the difficulty of the scenario is based on the performance of the employees. Due to privacy concerns, the difficulty is determined
per group of at least 10 employees, but this will still contribute in making the exercise a pleasant experience.

6. VERIFICATION AND VALIDATION
In accordance with Deloitte consultants, the design has been verified. Some discussion points remain, that all stem from the commercial interests of Deloitte as a company. For instance, Deloitte is hesitant to enforce some of the steps on clients, if the client is not convinced of the advices. The design has ethical values as a focus point, and is verified in accordance with the same Deloitte consultant as representing the values in an optimal way. The discussion point origin from the fact that striving for an ethically sound process is not the only interest of both Deloitte and its clients. Balancing the commercial interest with the ethical interest is a challenge that must be faced during each implementation process. The verified designed step-by-step guide is the most ethical way of implementing MTPA, so decision makers are at risk when deviating from the step-by-step guide. Providing all the underlying reasoning behind the step-by-step guide will help decision makers in making well-informed decisions.

7. DISCUSSION
7.1 Usefulness of VSD, VAP and Value Hierarchies
Besides the resulting design to implement MTPA, this thesis has also contributed to the development of methodologies that are aimed at taking values into account in the design process. This thesis has combined several frameworks: VSD, VAT and the notional of Value Hierarchies. The resulting steps in the design process (as shown in Table 1), have been perceived as useful and complete in this particular design process.

Yetim (2011) lists several challenges with respect to the incorporation of values into the design process. A couple of those are also experience in this thesis. For example, it was hard to decide on the starting point of the design process. Finally, it was decided that the norms expressed by the interviewees should functioned as a starting point, with the empty framework to help structure the interview. One can only wonder how the design would look like if another starting point would have been chosen, such as the values mentioned in literature.

Another challenged, related to the one mentioned above, was how to justify the selection of the values, norms and requirements. While the final Value Hierarchy has been verified, it is unclear whether a totally different Value Hierarchy would also lead to a verified design, supported by convincing argumentation. It remains a subjective process.

Future research should aim to provide more practical frameworks for designers interested in incorporating values in their work, to lower the threshold of doing so and to promote including values in engineering design processes. Furthermore, these research efforts should be combined with finding ways to make the process less subjective, for example by incorporating quantitative methods. This would also contribute to the acceptance of the framework by engineers.

7.2 Effectiveness of security awareness training
There is a growing body of research dedicated to finding methods on how to raise the security awareness of end-users. However, criticism is voiced as well. For example, the papers of Herley address the absence of economical rationality when investing in security awareness measures. He states that most ‘advice offers to shield them [the end-users] from the direct costs of attacks, but burdens them with far greater indirect costs in the form of effort.’ (Herley, 2009, p. 133), and that ‘[..] most security advice simply offers a poor cost-benefit trade-off to users and is rejected.’

This thesis has taken as a starting assumption that embedded training against targeted phishing is beneficial, which is argued for by looking into the trend of phishing in chapter 3. However, during the validation, economic reasoning was first taken into account. This let to discussions about how much to invest in embedded training, see for example the discussion about the necessity of adjusting the difficulty of the attacks to each group of users. This was reasoned from the point of view of Deloitte, but this discussion also need to be held within organizations. Future research could focus on cost-benefit analyses of implementing MTPA within companies.

8. REFERENCES


Figure 3: Upper layers of the Value Hierarchy

Legend

- Value
- Value
- Norm
Figure 4: Step-by-step guide to implement MTPA in an ethical way