

Modelling User-Phishing Interaction

Xun Dong, John A. Clark, and Jeremy Jacob

Department of Computer Science University of York, York, United Kingdom

xundong@cs.york.ac.uk, jac@cs.york.ac.uk, jeremy@cs.york.ac.uk

Abstract—To protect users from phishing attacks system designers and security professionals need to understand how users interact with those attacks and be able to predict users' behaviours in a given situation. In this paper we introduce the first model to visualise user-phishing interaction. We present a method to accurately describe users' perceptions in a uniform and compact manner. Within the context of this model we have investigated: what exact mismatches may occur between perception and reality in an attack; how to detect those mismatches; and why users fail to do so. Using this model we also identify where the security tools/indicators are lacking, suggest new aspects for security evaluation for the user interface, and provide guidance on effective anti-phishing user education.

Keywords—Phishing, User Interaction, Decision Making Model

I. INTRODUCTION

Personal information and authentication credentials can be obtained by attacking users as well as hardware/software. Hardening barriers to technological compromise has attracted a great deal of attention over many years. However, as the difficulty of technological compromise has increased, the computing system users, who clearly possess valuable information, have become more attractive targets.

Phishing attacks are the best-known recent examples of this trend. The term phishing describes the fraudulent acquisition, through deception, of sensitive personal information such as passwords and credit card details by masquerading as someone trustworthy with a real need for such information [16]. Financial losses stemming from phishing attacks have risen to more than \$3.2 billion with 3.6 million victims in 2007 in the US [12]. In the UK losses have doubled to \$46m in 2005, from \$24m in 2004, while 1 in 20 computer users claimed to have lost out to phishing in 2005.

If we are to make progress in reducing the damage caused by phishing attacks, we need to find out: how users decide what actions to take during the interaction with a phishing attack; and what are the factors that influence this decision making process. The answers to those questions can improve our understanding of why users fall victim to phishing attacks, and help security researchers to invent countermeasures that can effectively protect users. In the past few years, security researchers have made many useful discoveries regarding human factors in security, particularly in phishing. However, their findings vary in depth and there is a clear need to place such knowledge within a coherent framework. This will allow findings of research to be more easily absorbed and used as design

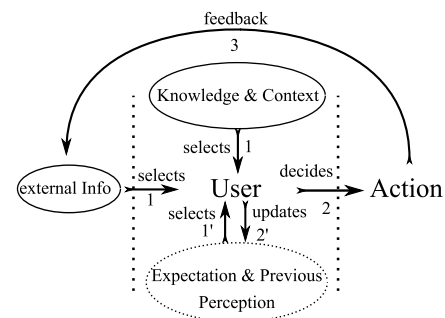


Fig. 1. The Overview of User-Phishing Interaction

guidance, as well as clarifying which areas need more research attention.

To tackle the issues identified above, we model the user-phishing interaction from the decision making point of view. Phishing attacks achieve their goals by making victims carry out actions which lead to the compromise of confidential information. The actions users take are the realization of the decisions they have made. Attacks are trying to manipulate victims to make certain decisions. Hence, how people make decisions when encountering phishing attacks is what really matters.

II. USER-PHISHING INTERACTION MODEL

Having reviewed a comprehensive range of related literature, we then performed a cognitive walkthrough on over 400 sample phishing attacks collected by APWG [2] and Millersmiles [1]. During the cognitive walkthrough we tried to identify all steps users take in their decision making process, what information users selected for their decision making, and the assumptions/expectation users have in each step. Our goal was to combine the gathered information about how users interact with the phishing attacks with the existing human factors results to abstract the user-phishing interaction model. The model starts when users encounter phishing attacks and finishes at the point where all actions have been taken.

A. Overview of The Interaction

Like any other types of human-computer interaction, the user's role in phishing attacks is to retrieve relevant information, translate the information into a series of actions, and then carry out those actions. The overview of the model is shown in Figure 1.

There are three obvious types of information users can use when encountering phishing attacks, and they are connected to users in Figure 1 by solid arrows. The external information is information retrieved from the user interface (includes the phishing emails/communication) as well as other sources (such as experts' advice). The context is the social context the user is currently in. It is the user's perception of the state of the world, comprising information on things like recent big news, what is happening around the user, the user's past behaviour, social network, etc. Knowledge and context are built up over time and precede the phishing interaction. External information must be specifically retrieved during the interaction. The information items displayed on the user interface are the most obvious and immediately available external information to users. As a result, they are always used in the interaction. External information from other sources is selected only when certain conditions occur such as users' suspicion.

It usually takes more than one step to reach the action that could lead to the disclosure of the information. For example, an email based phishing attack may require victims to read the phishing emails, click on embedded URL links and finally to give out confidential information at the linked phishing web site. For each completed action, users have expectations of what will happen next (the feedback from the system) based on their action and understanding of the system. This expectation, as well as the perception constructed in the previous steps, are carried forward when users decide whether to take the next action. Since these two types of information exist only during the lifetime of the interaction, we present them differently in figure 1. This decision making happens each time before an action has been taken. Thus Figure 1 includes a feedback arrow from action to external info. Among the five types of information, the attackers can directly affect only the information displayed by the user interface.

B. The Decision Making Model

There are two types of decisions that users make in user-phishing interaction; the first one is to decide a series of actions to take, and the second is to decide whether to take the next planned action or not. Users usually make the first decision consciously and the second subconsciously. As a result, user behaviour in the two types of decision making are not the same, and the first decision's outcome would also affect users' ability in detecting phishing attacks in the second decision making process. In this section we will describe the steps users take in both decision making processes.

By consulting decision making theories [3], [10], [11], [14], [15] and analysing the discovered phishing attack examples we find that both types of decision making during user-phishing interaction can be divided into the following three stages:

- construction of the perception of the situation;
- generation of possible actions to respond;
- generation of assessment criteria and choosing an action;

1) *Construction of the Perception*: The perception users have built in this stage is much more than just a goal. The perception is constructed by selecting the information

first, and then interpreting the information selected. The perception users construct can be described using the following four aspects:

- Space
 - Direction of the interaction;
 - Media, through which the communication and interaction occurred;
- Participant
 - Agent, who initiates the interaction;
 - Beneficiary, who benefits;
 - Instrument, who helps accomplish the action;
 - Object (often personal/business confidential information), the target of the action;
 - Recipient, who receives the action;
- Causality
 - Cause of the interaction;
 - Purpose of the interaction, this is usually the goal attackers want users to believe;
- Suggestion
 - Explicit suggested actions to reach the goal;
 - Implied actions to reach the goal;

In the rest of the paper we will use this method to describe user perception.

The mismatch between a user's perception and actual fact has been identified as the mismatch between user's mental model of the information system and actual implementation. This understanding is too general to be useful in analysis. Using the above four aspects we can discover that the false perception, which phishers try to engineer, has two mismatches with actual fact: 1) the perceived participant is not the actual participant; and 2) the perceived consequences are not the actual consequences.

These two mismatches exist in every phishing attack, because the real participants are the phishers, their phishing websites, etc. rather than the legitimate organisations or persons whom the victims trust; and the true causality is to steal people's confidential information rather than any causality phishers suggest. Failure to discover such mismatches allows phishing attacks to succeed. In a later section of the paper we will discuss how users could discover such mismatches and why they often don't.

2) *Generation of possible solutions*: In general factors such as time, knowledge, resource available, personality, capability, etc. all affect the set of actions one generates. Interestingly, analysis of existing phishing attacks suggests that the user's ability to generate possible actions is not one of the major factors in deciding whether they fall victim to phishing attacks or not.

In phishing attacks, victims do not generally need to work out a solution; the attacker kindly *provides* the victims with a "solution", which is also the action they want victims to take. For example, in email messages stating that there is a problem with a user's authentication data may also indicate that the problem can be "solved" by the user visiting a linked website to "confirm" his/her data. If the situation is as presented, then the "solution" provided is a rational course of action. Unfortunately, the situation is not as presented.

In some attacks the solution has not been explicitly given, but it can be easily worked out according to common

sense. For example, the attackers first send users an email appearing to be a notice of e-card sent by a friend. Such e-card websites are database driven, the URL link to access the card often contains parameters for the database search. The URL link that the attackers present to victims has two parts: the domain name of the website and the parameter to get the card. The former points to the legitimate website, but the parameter is faked. As a result, the user will see an error page automatically generated by the legitimate website. And at the end of the email attackers also present a link for people to retrieve the card if the given URL link is not working. This time the link points to the phishing website which will ask for people's address and other personal information. In this attack victims have not been told to click the spoofed URL link, but "common sense" suggests to use the backup link provided when the first link they have been given has failed.

We can even view phishing attacks as follows. The attacker tries to engineer a false perception within the victim's mind, and also tries to simplify the solution generation stage by telling the user what actions he/she should take to respond to the false perception. He or she must now decide only whether to take the suggested means of solving the problem – the user does not feel a need or desire to generate any alternatives. Simplifying users' decision making processes means users spend less time on selecting information. The chances of users discovering any misperception is also reduced.

This is one of the important difference between the user-phishing interaction and the general user-computer interaction where users have to work out the actions to reach the goals themselves. In fact, users need to do little at this stage, so we have not presented this stage in our graphical model illustrated later in this section.

3) *Generation of assessment criteria and choosing the solution:* To decide whether to follow the suggested action is the focus of this stage. A user generates the criteria to evaluate the resulting gains and losses of possible actions, then evaluates those actions and chooses the best one. Sometimes the criteria are already established such as one's worldview, personal preferences, etc. Sometimes criteria must be carefully developed. Each individual's experience, knowledge, personal preference and even emotional/physical state can affect the assessment criteria. However, most phishing attacks analysed didn't take advantage of the differences between users, instead they take advantage of what users have in common.

The suggested solution in a phishing attack is to satisfy some of the victim's assessment criteria and be rational according to the false perception engineered. For example, everyone wants their bank account authentication credentials to be secure, and so a solution which appears to increase security "ticks the box" – it simply appears to provide them with something they actually want. We shouldn't be surprised when they accept this solution and act on it. Similarly, users will (usually) feel a need to follow an organisation's policy, and so will most likely follow instructions that appears to come from authority. Being friendly, helpful to others, curious to interesting things, actions because of reciprocation (often used by social engineering attacks), are all common criteria.

As long as the victims have not discovered the mismatch between the false perception and the truth, they would indeed want to follow the suggested actions and the evaluation of this stage would most likely be "yes". Again this stage is also secondary in deciding whether users fall victim to phishing attacks, so we exclude it in our graphical model.

C. Graphical Model

In a user phishing interaction, the user first decides the sequence of actions, and then follows the decision making process concerned with whether to take the next planned action. The second stage decision making is carried out repeatedly prior to each action a user takes. Both types of decision making processes comprise the same three stages. However, their characteristics are different.

In the first decision making process, there are no expectations and no previously constructed perceptions constructed. Users select a wide range of external information to form their initial perceptions. In contrast, in the second decision making process, users have previously constructed perceptions and expectations to influence their current perceptions. The way they select the external information is also changed as well. The solution generation stage in the second type of decision making is minimised. The evaluation of whether to take the chosen solution (the next planned action) mainly depends on whether the feedback of previous action matches expectation, while the evaluation stage in the first decision making type is more general and flexible. For example when users click the hyperlinks embedded in phishing emails, they expect the web browser to present them the legitimate website that phishers try to impersonate. If they perceive the website presented by the browser (the feedback) as the legitimate one (the expectation), then they will carry out next planned action, which could simply be giving out their authentication credentials. If the users constructed a correct perception (they have been presented with a phishing website), which does not match their expectation, their next planned action will not be taken.

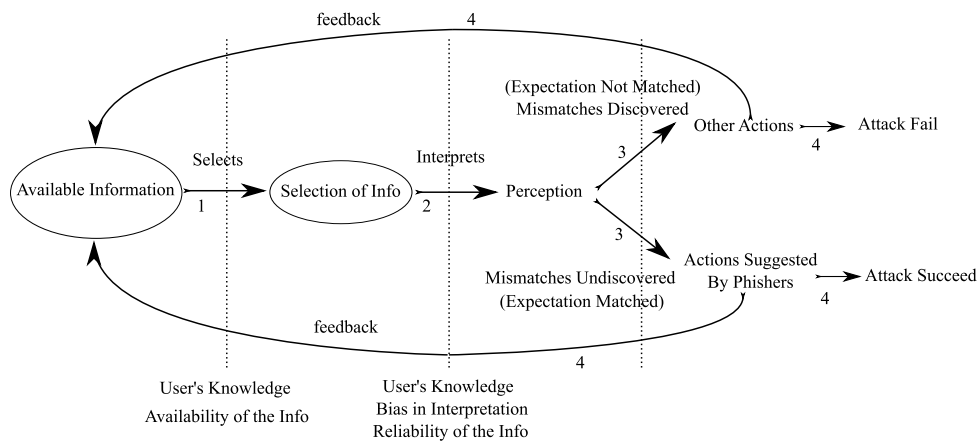
Regardless of the type of decision making process, construction of an accurate perception is key to detect phishing attacks. People's ability to work out solutions and evaluate the alternatives plays little part in preventing them from falling victim to phishing attacks in both types of decision making. Because the perception construction is so important, it is also the main focus of our graphical model, which is shown in Figure 2. The first cycle of this model describes the decision making regarding the planning of the sequence of actions to take, and the rest of the cycles are concerned with the deciding whether to take the next planned action or not.

III. FALSE PERCEPTIONS AND MISMATCHES

In this section we will further model the perception construction stage. We examine how those mismatches could be discovered, and why many users fail to do so.

A. How Mismatches Can Be Discovered

There are two types of mismatch: one concerns the participant, the other is causality. The discovery of the



Available information: External info, User's Knowledge, Context, Expectation and Previous Perception
 Note: the text underneath the vertical dashed line indicates the factors that affect the outcome of the corresponding step;

Fig. 2. The Decision Making Model

mismatch of participant (refer to previous section for how we define the participant) mainly relies on the information users select from the user interface. There are two types of information presented to users at the interface: 1) the meta data of the interaction; 2) the body of the interaction.

The meta data of the interaction can be used to form the space and participant aspects of the perception. For example, the URL of the website, digital certificates, sender's email address, etc. The body of the message is used for conveying the actual message of the interaction. It can be the text content of an email message, webpage visual content, a video or an audio clip.

The inconsistency between the body of the interaction and the meta data of the interaction can reveal the mismatch of the participant. For example the content of the phishing webpage suggests it is Ebay's website while the URL address suggests something else. The mismatches could be revealed if users selected both meta data and the body of the interaction to form the perception. Users with appropriate education could be expected to reliably discover those mismatches.

However, due to the vulnerabilities in the system implementation and design, the meta data could be spoofed to appear consistent with the body of the interaction. For example, the sender's email address could be configured to display any email address the attacker wishes; an example is illustrated in [8]. Rachna [4] has summarised a list of techniques phishers use to spoof the meta data. As a result, to discover the mismatch of the participants, users have to discover the inconsistency between the meta data and low level system data that is not presented to users at the interface. For example, if the attacker faked the sender's address, which has not been associated with a digital certificate, then users need to find out the original SMTP server who sent out this email and compare it with the server from where the email with the same email address is normally sent. Another example is when the phishing

website's URL address is spoofed ¹ to be identical with the legitimate site. Unless users can analyse the IP address of the phishing website as well as the IP address of the legitimate website, users cannot discover the mismatch. Other examples also include spoofing a phone caller's ID by using IP phone techniques.

Many people believe user education could solve the problem. But to discover the mismatches when meta-data is spoofed requires extra tools and knowledge which most users simply don't have and should not be expected to have. It is the system designers' responsibility to ensure information displayed on the user interface is resistant enough against most spoofing attacks, especially the meta-data. This is even more important for security tools/indicators, if they do not ensure their meta-data is reliable against spoofing attacks, they may invent new ways for phishers to engineer more convincing false perceptions.

Users need to be aware of certain contextual knowledge to be able to discover the mismatch of causality. The mismatch will be discovered if the contextual knowledge contradicts the body of the interaction. For example, suppose a Citi bank's customer has received an email, whose sender's address appears from customer service of the Citi Bank. This email asks the customer to login to Citi's online banking website to solve a security problem, otherwise the user's account will be suspended. If the customer knows that Citi will never send emails to their customers, then he/she can quickly realise this is a phishing email, and discard the email immediately. However, some sophisticated phishing attacks take advantage of the truth to serve their evil purpose. For example, to make the online credit/debit card payment more secure VISA has launched "verified by visa" scheme. In this scheme users will create personal messages and passwords for their cards, and when they pay for online purchases, users will be asked to provide the password as well as the card details. Phishers have taken advantage of this scheme and send users phishing

¹A phishing attack that targets Citi Bank customers has used this techniques. The story is published at <http://blog.washingtonpost.com>

emails which ask them to join this scheme, although links provided in emails are lead to phishing websites. In this case, it would be very difficult to discover the mismatch of causality unless users are aware of when VISA will send emails to its users and what emails have been sent. Unlike the mismatch of participant, the mismatch of causality is not always discoverable from the user side, as users can not be expected to possess the required contextual knowledge.

B. Why Users Form False Perceptions and Fail to Discover The Mismatches

Users do not solve the actual problem nor respond to the actual situation, they make decisions purely based on their perception [14]. Unsurprisingly, in phishing attacks victims invariably perceived the situation erroneously and solved the wrong problem. The victim's response is flawlessly rational according to the perception, he/she may execute an entirely cogent plan to react to the perceived situation. The problem is, of course, that this underpinning perception is simply wrong.

Users form a false perception and fail to discover the mismatches because:

- The selection of the information (especially meta data) is not sufficient to construct an accurate perception to reveal the mismatches;
- The information selected has not been interpreted correctly.
- They have inaccurate/incomplete expectations.
- Once engaged in the second type of decision making (regarding whether to take the next planned action) users often exhibit less suspicion (and less critical thinking generally).

1) *Insufficient Information*: There are three causes for insufficient selection of information:

- Meta data presented to users at the user interface is not sufficient;
- Users do not have the knowledge to allow them to select sufficient information;
- Users have not paid enough attention to security and hence some important information has not been selected.

Some user interfaces do not provide enough meta data information. For example, in mobile phone communication, the caller's ID can be hidden. In the UK, banks or other organisations often call their customers without displaying the phone number. The phishers can just call a customer to impersonate legitimate organisations by hiding their phone number as well. Even when the phone number is displayed, the users may still not be sure of the identity of the caller, because there is still no information (at the user interface) that can help them to decide who actually own a number. Here recognition of voice would not help because most likely the user will not have built up a relationship with any specific caller from that organisation (e.g a legitimate caller could be one of a considerable number in a call center.)

Some organisations and companies share users information, and users can login to their accounts by using any member's websites (UK's mobile phone retail companies: CarPhoneWarehouse, E2save.com and Talktalk.com share

their customer information, and users can use the same authentication credentials to access their accounts at any one of the three websites.). But how could a user tell whether a site belongs to a legitimate organisation or a phishing website by looking at the URL alone?

Insufficient information could also be the consequence of the meta information displayed on the interface being unreliable. As discussed in the previous section, if the meta data information can be spoofed by attackers, then users need more information to help them verify the meta information they get from the interface. Often such additional information is not provided to users at the interface.

Users do not pay enough attention to security related meta-data, because they are more interested in productivity. They want to solve the problem and react to the situation as efficiently as possible. Security related meta data seem unrelated to many users' primary goals. As a result meta data has often been ignored.

The other two causes have already been discussed thoroughly by Rachna Dhamija [4].

2) *Misinterpretation*: On the other hand, the way people interpret the information selected is not error-free. We summarised biases within users' interpretation of information [7], [9]:

- The existence of the HTTPS in the URL means the site is not a phishing website and should be trusted, HTTPS only indicates the use of TLS/SSL protocol, phishing websites can use it as well;
- The appearance of the padlock at the bottom of the browser or in the URL bar means that the web site visited is not a phishing website and should be trusted, the padlock only indicates secure communication and it has nothing to do with the identity of the website;
- The appearance of the digital certificate means the site is secure and should be trusted. The phishing website can have a digital certificate as well, it is the content of the digital certificates that matters not the appearance;
- The sender's email address is trustable. In reality it may be spoofed;
- The professional feeling of the page or email means it might come from the legitimate source;
- Personalisation indicates security;
- Two URL's whose semantic meanings are identical then the two sites are also identical, for example www.mybank.com equals www.my-bank.com
- The text of hyperlink indicates the destination of the link.
- Emails are very phishy, web pages a bit, phone calls are not.

Besides the misinterpretation, some users also may not be able to interpret the information presented to them at all. This is often the case for some security warnings or error codes. If people do not understand the information they receive, they will not be able to use it to form any meaningful perception and discover the mismatches.

3) *Inaccurate/Incomplete Expectation and Perception Ability Drop*: When users execute the planned sequence of actions, they inevitably have expectations on what the system will present to them or how the system should respond to their actions. Many users tend to select only the information that can be used to confirm their expectation.

Such selection bias has been discovered and proved in other situations [5], [13], [14], [17], but to what extent it affects the user's selection in phishing attacks is unknown.

Let's first analyse the result of a phishing attack study [6]. In this study participants have been categorised into two groups. The authors performed a controlled study on one group and the other group was unaware of the study whilst attacks were launched. In the controlled group, technology students all identified the phishing attacks, while in the other group over 36% of the technology students fell victim. The phishing website's domain name was www.whuffo.com which is significantly different from the legitimate website www.indiana.edu. However, the authors did not investigate the deeper reasons why the technology students performed so differently in two groups.

We believe this is because some participants of the first group have not got an accurate/complete expectation, and secondly their ability to select sufficient information to form a perception has been compromised by the previous perception constructed by reading the email sent to them by authors. Their expectation after clicking the link embedded in emails was that the legitimate website www.indiana.edu should be displayed. Those student victims may expect only the legitimate website's content, they have not expected the URL of the webpage is also consistent with the content of the page. Because they are biased to select the information that can be used to confirm their expectation, the URL displayed in the web browser is neglected. As a result, they fall victim to this attack.

To prove our point we carried out another experiment, and in this study we proved that technology students can detect such mismatches when they do not have such expectation created by reading the previous email. We first hosted a web page, whose appearance is identical to the home page of www.facebook.com, under the domain www.fakepage.biz. Like the pair of URLs in Jagatic's study, the mismatch of the two URLs is obvious. 19 computer science research students volunteered to take part in this study. They are all aware of phishing attacks. Along with another 9 legitimate websites we asked them to identify the phishing webpage among them. They all immediately discovered the mismatch.

Actually in Jagatic's study, all the technology students in controlled group successfully detected the attack as well. So why do the same students behave so differently for conducting the same exercise. Our interpretation is that in such a controlled condition students involved would be aware that anything given to them in the study is artificial and for research purposes. Hence they do not have the strong expectation as they would have in an uncontrolled environment. This may explain in general why participants in controlled users studies behave more securely than they usually would.

IV. CONCLUSION

In this paper we have introduced a user-phishing interaction model from a decision making point of view. This model is used to analyse how users can detect phishing attacks and discuss why users often fail to do so.

By researching phishing from this new angle, we have made some interesting discoveries. The result model sug-

gests that: a) System designers should focus more on providing security tools/indicators for communication channels. At the communication stage, systems have the chance to fully utilise users' perception ability, as users are in the initial stages of the first decision making process and are likely to pay more attention to security indicators and to be more suspicious. Currently most efforts have been focused on improving the usability of authentication on web pages where users decide whether to take the next planned action. However, in the second type of decision making, users' ability to form accurate perceptions is compromised. b) System designers should also pay more attention to evaluate how likely the information displayed on the security tools/indicators can be spoofed, and eliminate the vulnerabilities that could lead to such spoofing attacks; c) when educate users, rather than injecting a new step – detection into user's behaviour model, it would be more effective to focus on improving users' ability to form an accurate perception when interacting with computing systems by:

- teaching them the sufficient set of information they should select in different contexts;
- teaching them how to correctly interpret the information selected.
- training users to have more complete/accurate expectations.

REFERENCES

- [1] Millersmiles home page.
- [2] Anti-phishing work group home page, 2007. <http://www.antiphishing.org/>.
- [3] T. Betsch and S. Haberstroh, editors. *The routines of decision making*. Mahwah, NJ ; London : Lawrence Erlbaum Associates., 2005.
- [4] R. Dhamija, D. Tygar, and M. Hearst. Why phishing works. *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems.*, ACM Special Interest Group on Computer-Human Interaction:581–590, 2006.
- [5] A. H. Hastorf and H. Cantril. They saw a game: A case study. *Journal of Abnormal and Social Psychology*, 49:129–134, 1954.
- [6] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *ACM Communication*, October 2007.
- [7] M. Jakobsson. Human factors in phishing. *Privacy & Security of Consumer Information '07*, 2007.
- [8] M. Jakobsson and J. Ratkiewicz. Designing ethical phishing experiments: a study of (rot13) ronl query features. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 513–522, New York, NY, USA, 2006. ACM Press.
- [9] M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y. kyung Lim. What instills trust? a qualitative study of phishing. In *Extended abstract, USEC '07.*, 2007.
- [10] I. Janis and L. Mann. *Decision making: a psychological analysis of conflict, choice, and commitment*. Free Press, 1979.
- [11] G. Klein. *Source of Power: How people make decisions*. MIT Press, Cambridge, MA, 1998.
- [12] T. McCall. Gartner survey shows phishing attacks escalated in 2007; more than \$3 billion lost to these attacks. Technical report, Gartner Research, 2007.
- [13] D. McMillen, S. Smith, and E. Wells-Parker. The effects of alcohol, expectancy, and sensation seeking on driving risk taking. *Addictive Behaviours*, 14:477–483, 1989.
- [14] S. Plous. *The Psychology of Judgment and Decision Making*. Number ISBN-10: 0070504776. McGraw-Hill, 1993.
- [15] F. Schick. *Making choices : a recasting of decision theory*. New York : Cambridge University Press, 1997.
- [16] D. Watson, T. Holz, and S. Mueller. Know your enemy: Phishing. Technical report, The Honeynet Project & Research Alliance, 2005.
- [17] G. Wilson and D. Abrams. Effects of alcohol on social anxiety and physiological arousal: Cognitive versus pharmacological processes. *Cognitive Research and Therapy*, 1:195–210, 1977.