Exploiting curiosity and context:
How to make people click on a dangerous link despite their security awareness

By Zinaida Benenson[1], Freya Gassmann[2], Robert Landwirth[3]

Executive Summary

Messages containing links to malware-infected websites represent a serious threat. Despite the numerous user education efforts, people still click on suspicious links and attachments, while their motivations for clicking or not clicking remain hidden. We argue that knowing how people reason about their clicking behavior can help defenders in devising more effective protection mechanisms. To this end, we report the results of two user studies in which we sent to over 1600 university students an email or a Facebook message with a link from a non-existing person, claiming that the link leads to the pictures from the party last week. We registered the click rates, and later sent a questionnaire to the participants. The questionnaire first assessed their security awareness and then asked them about the reasons for their clicking behavior.

When addressed by first name, 56% of email and 38% of Facebook recipients clicked the provided link. When not addressed by first name, 20% of email and 42.5% of Facebook recipients clicked. While respondents reported a high awareness of the fact that clicking on a link can have negative consequences (78%), statistical analysis showed that such awareness is not connected to their reported clicking behavior.

By far the most frequent reason for clicking was curiosity about the pictures (34%), followed by participants explaining that the content or context of the message fits their current life situation (27%), such as actually having been at a party with unknown people last week. Moreover, 16% thought that they know the sender. The most frequent reason for not clicking was unknown sender (51%), followed by the explanation that the message does not fit the context of the user (36%).

Since the above mentioned reasons are easy to reproduce in fake messages (make a person curious about some piece of content behind a link, spoof a sender they might know or a life situation they might relate to), it should be possible to make virtually any person click on a link. Expecting from the users error-free decision making about clicking on a link or an attachment seems to be highly unrealistic, even if they are provided with effective awareness training.

Moreover, while sending employees fake spear phishing messages from spoofed colleagues and bosses may increase their security awareness, it is also quite likely to have negative consequences in an organization. People's work effectiveness may decrease, as they will have to be suspicious of practically every message they receive. This may also seriously hamper social relationships within the organization, promoting an atmosphere of distrust. Thus, organizations need to carefully assess the pros and cons of increasing security awareness against spear phishing. In the long run, relying on technical in-depth defense may be a better solution, and more research and evidence is needed to determine which level of defense non-expert users are able to achieve through security education and training.

[1] zinaida.beneneson@fau.de Friedrich-Alexander Universität Erlangen-Nürnberg (FAU), Germany
[2] f.gassmann@mx.uni-saarland.de Universität des Saarlandes, Germany
[3] robert.landwirth@fau.de Friedrich-Alexander Universität Erlangen-Nürnberg (FAU), Germany

1 Introduction

Social engineering attacks that persuade users to click on a malware-infected attachment or a link have become a standard means of gaining a first entry point into systems during APT (Advanced Persistent Threats) attacks and data breaches and have recently caused substantial damage in form of ransomware. Moreover, such attacks may be conducted not only via email but also via Facebook or Twitter, as many corporate users communicate via social media.

The popularity of this attack vector has inspired a lot of research on the susceptibility of the users. For example, researchers have shown that sending an email from a spoofed social network friend increased the success rate from 16% to 72% [1] and that emails with logos are significantly more difficult for the users to recognize as phish [2].

In security practice, sending fake phishing emails to employees has become a popular method of assessing their security awareness, with numerous commercial tools designed for this purpose [3,4]. A sophisticated tool for this kind of security analysis, comprising the analysis of social ties and social media communication between the employees, was presented at Black Hat 2015 [5].

Despite the numerous user education efforts, people still click on suspicious links and attachments and their motivations for clicking or not clicking remain hidden. We think that knowing how people reason about their clicking behavior can help the defenders in devising more effective protection mechanisms. In this article, we consider the following questions:

- Do people react to a “suspicious” link differently depending on whether the link was received via Facebook or via email?
- Which factors are related to the success rate of the attack? Examples of possible factors are gender of the sender and the receiver, information on sender's Facebook profile, Facebook friend request accompanying the message, general security awareness or personal experience with malware.
- What are the reasons for clicking or not clicking on the link?
- Is “pentesting the humans” by sending them messages with “phishy” links a good idea? What are the ethical and practical challenges of this popular method of raising security awareness?
- To which extent can we “patch” the humans? In other words, what is the feasible level of defense that non-expert users can achieve through education and training?

To answer these questions, we report the results of two user studies where we sent an email or a personal Facebook message with a link from a non-existing person to over 1600 users (mostly university students), claiming that the link leads to the pictures from a party. When clicked, the corresponding webpage showed an “access denied” message. We registered the click rates, and later sent a questionnaire to the participants which first assessed their security awareness and then informed them about the experiment and asked about the reasons for their clicking behavior. We also asked how this experiment is going to affect their future security decision making and how they felt about participating in the experiment without knowing that they were going to be phished.

2 Study design: That’s ethics, stupid

When experimenting with people, it is always important to consider ethical implications of the experiment, i.e., whether the participants are likely to be harmed. Although clicking on a link is unlikely to result in physical harm, it can (and does) result in psychological harm, such as worrying (“did I infect my computer?” [8]) or sometimes even monetary harm if people worry so much that they decide to let somebody check their computer for malware (we had one instance of this in another experiment). Still, this harm is unlikely to be greater than the harm from a real spear

phishing message. Moreover, as this type of study promises to find out more about susceptibility factors and possible defenses, their benefits are usually considered to overweigh their harms.

An important question is, whether it is ethically permissible to conduct phishing studies without participants' consent, i.e. to just send to them “phishy” messages and register the clicking rates. We strongly recommend explicitly recruiting participants for any phishing study, while explaining all study details may not be necessary. Especially organizations should be very careful here, as “pentesting” participants without their consent can backfire severely. People do not like being treated as guinea pigs and may become disgruntled as a result.

For our study, we recruited users by asking them to participate in a survey about their “Internet habits” and debriefed them after their participation by sending them a message containing cumulative anonymized statistics about the study results and an explanation why clicking on a link might result in a security incident.

We also encountered an additional restriction in our study design. The initial idea was to (1) recruit people for filling out a questionnaire about their “Internet habits”, (2) send an individualized phishing link to the recruited people and (3) send them an individualized questionnaire link about their link clicking behavior (and not their Internet habits) afterwards. Thus, we wanted to connect, for each person, their link clicking behavior with their questionnaire answers.

This study design has a serious flaw: usually, questionnaires should be filled in anonymously, because otherwise, participants might not express their real, unedited opinions. Unfortunately, our initial study design idea would have effectively de-anonymized the survey answers, as it would be possible to connect the previously sent individual clicking links with the answers and, ultimately, with the email addresses of the participants. Therefore, we had to change the design. In our final design, we sent a truly anonymous questionnaire to the participants. This questionnaire first assessed the security awareness of participants (for example, can link clicking be dangerous?), then informed them about their inadvertent participation in the experimentand asked whether they clicked or not as well as the reasons behind their clicking, the latter in the free-text form.

3 Study 1: Party Last Week

Study 1 was conducted in September 2013. We recruited 240 Facebook (120 male, 120 female) and 158 email (123 male, 35 female) participants from our university. They were sent the following message:

Hey *<receiver’s first name>*,
here are the pictures from the last week:
http://*<IP address of our server>*/photocloud/page.php?h=*<USER ID >*
Please do not share them with people who have not been there :-)
See you next time!
<first name of the sender>

The message was targeted at our audience, but not too carefully. On the one hand, it used, as the name of the sender, names out of the top 10 German first names and family names in the generation of our receivers (university students in their twenties), such as Sabrina Müller and Frank Bauer. Further, it referred to this indispensable attribute of the student life: a party. On the other hand, the message was in plain text (not HTML), and the link contained a bare IP address, which has been known to alert users to some extent [9]. On Facebook, we used “open” sender profiles which contained some pictures of the sender, his/her timeline etc. and “closed” sender profiles with minimum information and no pictures., We also sent a friend request to the half of Facebook

participants, in order to see if this influences the click rates. Three weeks after sending the messages we sent the questionnaire to the participants.

56% of email recipients and 38% of Facebook recipients clicked on the link. Most first clicks occurred in the first 12 hours after the message was sent and none of the following made statistically significant differences for the success rate: gender of sender, gender of receiver, Facebook friend request or amount of information on sender's Facebook profile.

When analyzing the questionnaire results, we met with an unpleasant surprise: Only 20% of participants said that they clicked on the link, with answers of an additional 15% not known because they did not participate in the questionnaire. Because this was much lower than the number of users that actually clicked (45%), we decided not to analyze questionnaire results at all and conduct a second and improved study instead.

4 Study 2: New Year's Eve Party

To solve the aforementioned problem of credibility we decided, for our second study, to send different survey links to the participants that received the suspicious link by email and to the participants that received the suspicious link on Facebook. In this case we can verify, at least partially, the credibility of survey answers by asking the participants on which of the two communication channels they received the link. Another possible solution would be to send a different survey link to the clickers as to the non-clickers. Unfortunately, in this case it is impossible to argue with absolute certainty that each particular person would belong to a big enough anonymity set, as we cannot know the clicking rates beforehand.

If we assume that people were not lying in Study 1 but truly could not remember clicking on the link as some evidence suggests [8], the quality of the responses could be improved by sending the survey shortly after the link clicking event. We decided to send the survey 24 hours after participants clicked on the link. An alternative, namely sending the survey immediately after the first clicking event, was also debated and rejected. We were afraid that participants might feel that they were being tracked, which would result in negative emotions. Non-clickers received the survey 7 days after the message was sent.

We sent a very similar message to the 975 email participants (710 female, 265 male) and 280 Facebook participants (200 female, 80 male) this time, but as the second study was conducted in January 2014, we particularly mentioned the New Year's Eve party. Moreover, we were surprised that in Study 1 email success rate was significantly higher than the Facebook success rate, as our intuition was reverse. In agreement with recent findings [10] we came to the conclusion that addressing the participants by first name might have made the difference and decided to drop this element of the message. This time, 20% of email and 42% of Facebook recipients clicked. As in the Study 1, no other factors were correlating significantly with success rate (gender, friend request, Facebook profile).

We utilized the user agent parser ua-parser [11], more specifically its Python implementation by Google [12], to extract browser, operating system, and device data from user agent strings stored in our Web server logs. By doing so, page requests related to bots or spiders, such as Facebook or Google, were removed. Most participants (69%) executed the first click form a Windows PC, followed by Android (12%) and iOS (11%) devices. The most popular browser was Firefox (41%), followed by Chrome (18%) and Internet Explorer (10%). There were no significant differences between email and Facebook participants. 27% of women and 15% of men executed the first click from a smartphone.

The maximum number of times a single user clicked on our link was 24, followed by 14 and 9. 55% of the participants clicked only once, 24% clicked twice, and 10% clicked three times. 30% of all registered clicks via email and 44% on Facebook happened in the first 2 hours after sending the message. After the first 24 hours, the percentage of individual clicks reached 72% on both media. 95% of all clicks were registered during the first 5 days.

4.1 Reliability of the survey results

Almost all survey participants could remember accurately whether they received the message via email or via Facebook. Whereas email participants reported a very accurate clicking rate (20% clicked, 16% reported that they clicked), the reported clicking rate of Facebook users was much lower than our factual clicking data: 42.5% versus 16%. Finding an explanation for this effect seems to be difficult. We do not have any reason to assume that the Facebook participants lied more than the email participants or have worse memories. Maybe the difference lies in the use of Facebook and email as communication media. From the perspective of day to day behavior of users, handling messages on Facebook might be different from handling emails. Users might use these media at different times of the day, with different mind sets, different goals and also different reaction and action patterns and thus leaving room for the medium effecting a user's behavior or their retention of this behavior.

4.2 Survey: Attack success factors

Respondents of the survey reported a high awareness of the fact that clicking on a link can have bad consequences (82%). However, statistical analysis showed that this was not connected to their reported clicking behavior. Further, there was also no connection found to them having been on a party shortly before the message was received. People who claim that they are good at recognizing messages from criminal senders also less likely reported that they clicked on the link. Moreover, people that reported having paid a lot of attention to the message and having been in positive mood at the moment of its reception more likely reported that they clicked. The latter could indicate that the participants were in the state of "cognitive ease" when processing the message, which demonstrably leads to taking mental shortcuts in decision-making processes [13].

4.3 Reasons for clicking

By far the most frequent reason for clicking was curiosity, stated by 34% of reported clickers. These participants knew that the pictures cannot be for them, but were interested in the supposedly funny or private content. 27% of clickers explained that they could identify with the situational context given in the message, such as actually having been at a New Year's Eve party with unknown people. Moreover, 16% thought that they know the sender.

4.4 Reasons for non-clicking

The most prominent reason for not clicking was that participants didn't recognize the sender's name (i.e., as the name of an acquaintance). This reason was given by 51% of non-clickers. Although the sender's name is an important indicator to identify scam messages, only three users explicitly commented that one cannot fully rely on this criterion alone, as dangerous messages can also be sent by known senders. Quite a lot of people stated that the message did not fit their New Year's Eve celebration (36%) or their expectations (12%), for example in how the message was formulated (regarding choice of words or other matters of style). Interestingly, 6% of non-clickers explained that they thought that the message was genuine, but did not click because they wanted to protect the privacy of the sender or were just not interested. That is, these people were protected from the would-be threat by their decency or the lack of curiosity.

4.5 Reported impact of the study

We asked the participants to indicate whether their participation in the study is going to have an impact on their future behavior, and if yes, to explain how. 396 respondents said that the study is not going to have any impact, mostly stating as a reason that they are already behaving very carefully,

especially with the messages from the unknown senders. Out of the 191 participants that reported that the study is going to have an impact on them, 31% stated they are going to be more careful in the future without mentioning any specific changes in their behavior, 28% explained that they are going to be careful with the messages and/or links of unknown senders or that they are going to think more about risks of the Internet usage (26%).

62% of the respondents indicated a positive attitude towards their participation in the study, while 24% had a neutral attitude, and the rest (13%) reported a negative attitude, with clickers being more likely to give positive ratings. When asked whether such studies should be conducted in the future, 85% said “yes”, 13% said “not sure” and 2% said “no”.

5 Lessons learnt

Our study and its specific message design revealed susceptibilities to scam in some people as well as the reasons behind their susceptibility, but I think that the lesson learned is broader. With careful design and timing of the message, it should be possible to make virtually any person click on a link, as any person will be curious about something, interested in some topic, or find themselves in a life situation that fits the message content and context. Thus, expecting users to attain error-free decision making seems to be highly unrealistic, even if they are provided with effective awareness training.

For example, if a person’s job requires them to process a lot of invoices sent to them by email, they are also going to click on a ransomware-infected file called “invoice”, as this fits their job expectations [14]. And if we teach these people to be careful with invoices, they will start missing the real ones, and they are not going to be appreciated for this. However, most likely they will disregard the education attempts, because the only way for them to get their job done in time is to process their emails as quickly as possible, without wasting time with extra security checks.

Moreover, trying to involve users into perimeter defense by means of catching them on dangerous link clicking might have unintended negative consequences. For example, employees of an organization may become disgruntled and unmotivated if they find out that they are being attacked by their own security staff [15]. Moreover, sending employees messages from spoofed colleagues, friends and bosses, although it might raise their security awareness, may also seriously hamper their work effectiveness and efficiency and damage social relationships within the organization, promoting an atmosphere of distrust. Being suspicious of every message that contains typographical errors because it might have been sent in a hurry from a mobile device, or is otherwise a bit strange, will deprive people from (usually reliable) decision heuristics such as “this message fits my current expectations” or “I know the sender”, making them less efficient in their jobs, if these jobs require processing of a high number of messages.

On the whole, more research and evidence is needed to determine the feasible level of defense that the non-expert users are able to achieve through security education and training, and it seems to be clear that achieving a 100% defense rate is not only impossible but also economically unjustifiable. This further implies that the protection of the users and organizations against such threats should rely on in-depth, rather than perimeter defense [16], as the attackers will almost unavoidably get in. Trust and social relationships, decisional heuristics that make our life easier, and also natural and creative human traits such as curiosity, will remain exploitable forever, as humans (hopefully) cannot be patched against these exploits.

6 References

[1] Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. Communications of the ACM, 50(10), 94-100.
[2] Blythe, M., Petrie, H., & Clark, J. A. (2011, May). F for fake: four studies on how we fall for phish. In: SIGCHI Conference on Human Factors in Computing Systems (pp. 3469-3478). ACM.
[3] http://phishme.com
[4] https://www.knowbe4.com/phish-alert
[5] http://www.avasecure.com
[6] (in German) http://www.heise.de/security/meldung/Phishing-Test-bei-der-Berliner-Polizei-3028064.html
[7] (in German) http://www.tagesspiegel.de/berlin/telefonstreich-bei-der-berliner-polizei-attacke-noch-nicht-aufgeklaert/12659326.html
[8] Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going spear phishing: Exploring embedded training and awareness. Security & Privacy, IEEE, 12(1), 28-38.
[9] Onarlioglu, K., Yilmaz, U. O., Kirda, E., & Balzarotti, D. (2012, January). Insights into User Behavior in Dealing with Internet Attacks. In NDSS.
[10] http://arstechnica.com/security/2016/04/crypto-ransomware-targets-called-by-name-in-spear-phishing-blast
[11] T. Langel. ua-parser. https://github.com/tobie/ua-parser
[12] Google Inc. uap-python. https://github.com/ua-parser/uap-python
[13] Kahneman, D. (2012). Thinking, fast and slow. Penguin Books.
[14] http://arstechnica.com/security/2016/02/locky-crypto-ransomware-rides-in-on-malicious-word-document-macro
[15] Sasse, A. (2015). Scaring and Bullying People into Security Won't Work. IEEE Security & Privacy, (3), 80-83.
[16] Straight, R. J. Whatever You're Doing Isn't Good Enough: Paradigm Shift in Approach to Cybersecurity Needed to Minimize Exposure. Fintech Law Report. Sep/Oct 2014.