[31] Dissecting Social Engineering Attacks Through the Lenses of Cognition

Pavlo Burda (Eindhoven University of Technology), Luca Allodi (Eindhoven University of Technology) and Nicola Zannone (Eindhoven University of Technology).

Reviews Phase 1

Review 1

Time: Jun 25, 14:52 [MODIFIED AFTER REBUTTAL]

Overall evaluation: 2 (accept)

Rebuttal --- The changes proposed in the rebuttal letter address my concerns. --- ===== Brief paper summary =====

The paper presents a framework for the analysis of SE attacks by fitting its characteristics in defined blocks (stimulus, parameters, perception, attention, anomalies, etc.). It then instantiates the framework with 4 examples (two academic experiments, two real cases).

===== Strengths =====

- It covers a hot topic in which we lack a systematic approach

-The framework can be a first step towards a more systematic approach to compare experiments on attacks that make use of this kind of attack vector.

-The instantiation of the examples is extensive and covered in detail.

===== Weaknesses =====

-I see a straightforward application into empirical experiments on SE but I have more doubts about the application on real attacks. See detailed comment below.

===== Detailed comments for the author(s) =====

-Application of the framework to existing attacks: the authors propose to use the framework to "sheds light on the attacker capabilities [...] which allows to reason on the possible identity of the threat actor". I would be careful with these claims. Apart from splitting between a multi-step SE campaign from a one-shot one, the framework does not provide anything new that can help in identifying the capabilities and thus the identity. For example, some campaigns of the alleged state-sponsored Fancy Bear group employed a very simplistic SE attack in which they triggered users to click a link on an email related to the reset of their Gmail password (https://www.forbes.com/sites/kevinmurnane/2016/10/21/how-john-podestas-emails-were-hacked-and-how-to-prevent-it-from-happening-to-you/, https://citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/). The attacker's capability and identity cannot be obtained from the analysis of the SE attack as this is no different than many other SE attacks. The capabilities and identity are obtained from the analysis of the steps they performed once they got access, possible overlap in the infrastructures employed, etc.

===== Minor comments for the author(s) =====

-Fig.1: attack parameters and attributes should reflect into the stimulus. The authors could consider adding an arrow from the first to the latter. For example, if the attacker wants to target the Director of a company, he probably employs a certain stimulus (e.g. email) with certain attributes (e.g. a certain cognitive bias on possible collaboration) and not another (e.g. job offer on Linkedin).

-Fig. 2,3,4 missing the medium.

-Example of ‘tailored phishing against organizations’: As done for all other examples, I would add the explanation of the pretext utilized (i.e. booking holidays) at the beginning of the subsection. It will allow the reader to follow the paragraph more easily.

-Example ‘NGO spear-phishing’ and Fig.4: The (alpha_2)^s is the date when the email is sent and not the conference date. The date should be 04/03/2013 and not 04/03/2019.

Reviewer's confidence: 3 (medium)

Review 2

Time: Jun 10, 03:20

Overall evaluation: 1 (weak accept)

The paper presents a framework to analyze the cognitive process involved in a social engineering (SE) attack. The framework provides a way to break down a social engineering attack case into several cognitive processing steps from the victim’s perspective (such as perception, attention, and elaboration, behaviors) and describe how different factors (such as stimulus and attacker/victim characteristics) affect these steps. The authors provide 4 case studies on SE analysis using the framework.

Overall, I enjoyed reading this paper, especially the discussion of the building blocks based on cognitive science literature (Sec 2). The authors also analyzed more than one case study to show the framework is generally applicable.

My only concern is it is unclear what new insights the framework can provide by analyzing these SE cases. For instance, the authors mention that certain attacker strategies are better at triggering cognitive biases from the victims, and the framework explains why certain attacks are effective or ineffective. I don't find these "insights" to be very concrete, probably because this paper does not have any experimental studies. It appears the framework is useful to raise hypotheses --- it would be helpful if the authors can run validation experiments for these hypotheses too (maybe out of the scope of this paper).

Reviewer's confidence: 2 (low)

Rebuttal Letter

We thank both reviewers for the positive feedback and for their insightful comments and suggestions to improve the quality of our work.

The comments highlight two aspects of primary importance: 1) lack of clarity on claims of applicability of the framework for attacker identification, and 2) need for additional experimental work.

1. We agree completely with the first reviewer and will clarify in the revision of the paper that attacker identification requires an investigation of the attack's steps and of the infrastructure employed by the attacker, whereas a framing of adopted SE techniques may often not provide useful details to that goal.

2. We also agree with the second reviewer that experimental validation of the proposed framework is not in scope for this paper. Our contribution aims at making a first step in identifying a structure to bridge the gap between human cognition and SE effects. Our hope is that this work will spark new

experimental work in the community to validate and revise the framework where needed. We do plan to conduct some of that experimental work ourselves.

Reviewer 1 also had other minor comments on figures and clarity: we will address those for the camera ready. Thank you.

Reviews Phase 2

Review 3

Time: Jul 01, 14:05

Overall evaluation: 2 (accept)

This is an interesting submission that offers a framework to map the cognitive processes involved in social engineering attacks, a prevalent form of online risk. The submission summarizes the psychology and neuroscience literature and applies it to a cybersecurity problem. I think that this alone makes it a valuable contribution that will pave the way for more robust lab and field experiments. I would encourage the authors to pay more attention to the attackers' perspective: attackers possess a lot of implicit knowledge on what works and what does not in terms of SE effectiveness and this knowledge might help the authors think about which cognitive processes are the most important and which ones are more peripheral. In that respect, I would point to them the work of Kevin Steinmetz, a criminologist who has interviewed social engineers to understand how they see their victims and manipulate them (see for example his latest paper in Computers in Human Behavior: Performing social engineering: A qualitative study of information security deceptions). The interaction between cognitive and social processes seems like a natural avenue to deploy this framework further. The response to the first round of reviews seems appropriate to me.

Reviewer's confidence: 3 (medium)