Analysing Simulated Phishing Campaigns for Staff

Melanie Volkamer[1], Martina Angela Sasse[2], and Franziska Boehm[1]

[1] Karlsruhe Institute of Technology, Germany, {name.surname}@kit.edu
[2] Ruhr-Universität Bochum, Germany, martina.sasse@rub.de

Abstract. In an attempt to stop phishing attacks, an increasing number of organisations run Simulated Phishing Campaigns to train their staff not to click on suspicious links. Organisations can buy toolkits to craft and run their own campaigns, or hire a specialist company to provide such campaigns as a service. To what extent this activity reduces the vulnerability of an organisation to such attacks is debated in both the research and practitioner communities, but an increasing number of organisations do it because it seems common practice, and are convinced by vendors' claims about the reduction in clickrates that can be achieved. But most are not aware that effective security is not just about reducing clickrates for simulated phishing messages, that there are many different ways of running such campaigns, and that there are security, legal, and trust issues associated with those choices. The goal of this paper is to equip organisational decision makers with tools for making those decisions. A closer examination of costs and benefits of the choice reveals that it may be possible to run a legally compliant campaign, but that it is costly and time-consuming. Additionally, the impact of Simulated Phishing Campaigns on employees' self-efficacy and trust in the organisation may negatively affect other organisational goals. We conclude that for many organisations, a joined-up approach of (1) improving technical security measures, (2) introducing and establishing adequate security incident reporting, and (3) increasing staff awareness through other means may deliver better protection at lower cost.

Keywords: Social Engineering, Phishing, Security Awareness

1 Introduction

Although phishing attacks are not a new phenomenon, they are still a major threat to many organisations: small or large, national or international, public or private sector. There are a number of definitions of what a phishing attack is; in this paper, a broad definition. Phishers try to

- either steal the (digital) credential of their victims to harm them directly, or
- use stolen (digital) credentials to carry out attacks on others, or
- install malware on the victim's system, that can then be used to steal credentials or other information, or make files unaccessible and extort payments to have them restored.

A single employee who falls for a phishing attack can cause significant damage to an organisation, for instance if files that are needed for daily business are no longer

available. Sometimes, it can also be the starting point for further attacks on customers or suppliers.

To prevent their staff falling victims to phishing attacks, organisations resort to running simulated phishing campaigns. In a simulated phishing campaign, email messages with malicious links or attachments are sent to staff to see if they are 'vulnerable' to this form of attack, and then present those who are with education or training measures that aim to help them recognise this form of attack - and thus not fall for them in again. Given the plethora of security vendors offering toolkits or service for simulated phishing campaigns, many organisations are under the impression that this is an essential measure to defend against such attacks. In this paper, we will examine the different objectives and forms of phishing campaigns, and point out the challenges associated with with conducting them in practice; we also present the associated costs and potential side effects organisations should consider before deciding whether to implement such a campaign.

We first present various elements and types of phishing attacks (Section 2). In Section 3, we examine the objectives an organisation has when conducting simulated phishing campaigns. Section 4 presents the different forms of phishing campaigns, and what choices an organisation can make when implementing them. In Section 5, we examine phishing campaigns from security, legal, and human perspectives, to highlight the side effects and longer-term consequences for an organisation. Finally, in Section 6 we discuss to what extent the data collected during simulated phishing campaigns are a valid reflection of how vulnerable, or not, an organisation is to this form of attack.

2 Different forms and types of phishing messages

In this chapter we define the types of phishing messages attacker send[3]. Phishing messages can be sent via different channels, be it by email, via messages and/or posts in social media or social networks, via direct messages in messengers, or as text message. It is important for organisations to be aware that attackers increasingly use other channels - such as text messages - to trick staff into make contact with the attacker, and that in addition to malicious links or attachments, attackers may use media files, for instance voicemails that seem to come from the chief executive officer (CEO).

The contents of phishing messages can be dangerous in different ways. In a phishing message, the recipient is usually asked to perform one of the following actions:

1. disclose sensitive data such as access credentials, confidential documents, or credit card data,
2. transfer money or make calls, e.g. to supposed friends or business partners (e.g. in connection with the so-called CEO fraud),
3. disable or circumvent security measures, e.g. deactivate virus protection or install a (malicious "update",
4. click on links or go to go websites, which

[3] Some phishing messages are also referred to as spam. Spam includes any kind of unsolicited messages, so phishing messages are a sub-set of spam messages - and indeed many staff do not distinguish and use the "spam messages" instead of phishing

(a) either lead to a genuine-looking but fraudulent website, where sensitive data such as login details have to be entered, or
(b) lead to a website that attempts to install and distribute malware on your devices (just clicking on the link can cause immediate damage), or
(c) open dangerous attachments that contain malware or dangerous links.

The sophistication of phishing messages - and thus the difficulty involved in detecting them - varies considerably:

- Very easy to identify phishing messages contain noticeable spelling and grammatical errors[4] and/or incorrect presentation.
- Phishing messages that are moderately difficult to identify may look credible in terms of content and presentation, but may come from an implausible sender (e.g. sender's email address[5] or sender's phone number). In some email clients, only the sender's name may be displayed, and checking the sender's address requires an extra action from the user - e.g. to hover the mouse over the name.
- Phishing messages that are difficult to identify are plausible in terms of content, presentation, and sender. Accordingly, depending on the type of response desired by the phisher, the message can only be recognised by the account or telephone number, the URL behind the link, or the attachment type. Such messages can be sent because, for example, real message content is copied from large providers (so-called clone phishing), the email address is spoofed (faked), the salutatory address is replaced accordingly, and the corresponding information is exchanged.
- Phishing messages are very difficult to identify if the phisher has access to a genuine email account, and uses that to send plausible-sounding phishing messages - and sometimes even referring to a previous email communication. The email account is usually that of a person, e.g. a colleague, or another employee of a customer or supplier. Attackers have also managed to gain access to the email accounts of genuine service providers and sent phishing messages from there.

Phishing messages lead the recipient to believe that there is a - more or less plausible - reason why she should carry out the requested action. Attackers often add psychological triggers - such as creating time pressure, threatening punishment or promising gains - are used[6][7]. The triggers steer the recipient towards carrying out the action, and away from checking for signs that of a phishing message.

[4] Note that on the other hand, not all message with incorrect spelling and grammatical errors are phishing messages - with increasing digitalization and widespread use of social networking, and increasing awareness of conditions such as dyslexia, much non-malicious written communication contains such errors; when attackers impersonate some senders, it can even be interpreted as a sign of authenticity.

[5] It is important to check the sender's e-mail address and not just rely on the sender's name, because the latter is very easy to alter.

[6] Stajano,Wilson: Understanding Scam Victims: Seven principles for system security. Communications of the ACM 2011, 54(3):70-75

[7] Again, the same tactics are used by senders of legitimate messages .

Attackers pursue a range of different strategies: They either try to reach as many potential victims as possible with the same message, or they target their message at a specific person: In the case of *'classic' phishing*, the attacker sends the same message to all recipients available to him (i.e. not just to one organisation). Usually, the salutatory address is 'Dear customer, dear ladies and gentlemen'. The message is personalised only if this can be done automatically, e.g. because the attacker attempts to derive the name from the sender's information (such as the email address), or because the name or the gender are known in addition to the email address (e.g. because this information is also available on websites and can be read automatically). From the attacker's point of view, such a phishing attack is successful even if not all recipients react to the message, but only a few to whom the message appears plausible at the time - e.g. if a (phishing) message from Amazon one day after having placed on order. Classic phishing is mainly based on phishing messages with dangerous links and attachments, as these are likely to be clicked / opened by many recipients. Phishing messages of varying degrees of simplicity or difficulty are used in case of 'classic' phishing. *Spear phishing* is a form of phishing where attackers specifically attack an organisation or even a person. The attackers first collect information - either purely via the information freely available on the Internet about the organisation (e.g. customers, service providers, cooperation partners, or newsletter), or via the staff or even additionally via phone calls. Based on this bunch of information, organisation-specific phishing messages (e.g. from a customer, service provider, or cooperation partner) are then written. Due to the message's reference to the organisation, and possibly to one's own position and function in the organisation, spear phishing is generally much more difficult to identify than classic phishing.

3 Objectives of simulated phishing campaigns

In conducting a simulated phishing campaign, an organisation may pursue one or more of the following objectives:

Objective 1: To determine how vulnerable – or resistant – the organisation currently is to phishing attacks (and how many identified phishing attacks are being reported). Security staff may do this to obtain more budget for IT/information security and/or data protection activities, or to make the case that a mandatory security awareness campaign and/or security training should be introduced.

Objective 2: To demonstrate to staff who click that they vulnerable, and create a so-called "teachable moment". Here it is assumed that someone who falls for a (simulated) phishing message is particularly receptive to security awareness training immediately afterwards. At the exact moment when they recognise they might have potentially fallen victim to such an attack, staff are presented with information on how to recognise phishing messages, and on how to report them. Creating a "teachable moment" - as opposed to providing this information as part of general awareness campaigns or training - is supposed to be more effective because the awareness that one is vul-

nerable is supposed to focus attention and increase motivation[8]. The security training delivered at this point can be optional or mandatory. There are two types of settings:

- (objective 2a) In this case, the number of staff who fell for messages, or reported a messages, are not collected and reported. In this setting, the purpose of the simulated phishing messages is purely to raise awareness of the organisation's vulnerability.
- (objective 2b) In this case, the number of staff who fell for the message or reported a message are counted to evaluate the security awareness measure, and (hopefully) show that the campaign has decreased the organisation's level of vulnerability. This is typically measured as a percentage how many simulated messages resulted in a link or attachment are being clicked on.

Objective 3: (Scientific) evaluation of a security awareness measure deployed by the organisation (or parts of it) - e.g. an awareness campaign or training module. In this case, the simulated phishing campaign serves only to evaluate the effectiveness of the security awareness measure - which may be a new product, or a new security awareness measure created by researchers. Such quasi-experimental evaluations are often limited to subsets of staff.

Some organisations conduct campaigns purely for compliance reasons simply to be able to report to an auditor or regulatory authority that the organisation has "run awareness campaigns" or "trained staff". Indeed, some audit procedures do not ask for more evidence than that. However questionable, there is sometimes also the assumption that by virtue of having conducted a simulated phishing campaign, the organisation can 'offload' the responsibility in case of a breach on the staff member who fell for an attack, despite "having been made aware" of the risks. Since this is not a responsible approach, we will not consider this objective in this paper.

4 Simulated Phishing campaign designs

Simulated phishing campaigns involve sending various fraudulent messages to the staff of the organisation over a certain period of time. There are different ways of designing and conducting such campaigns .

As outlined in Section 1, phishing campaigns can cover different message channels, different types of dangerous content, different levels of difficulty of phishing attacks, with or without the use of psychological triggers, and different attack strategies. When we refer to a simulated targeted spear phishing campaign, we use that term "spear phishing campaign", and "phishing campaign" for general ones. In addition, different types of message content and sender type (e.g. the message comes from a person or organisation) can be used. Messages can be sent with or without reference to recent

[8] Kumaraguru, Sheng, Acquisti, Cranor, Hong:Teaching Johnny not to fall for phish, ACM Transactions on Internet Technology 2010, 10 (2):1-31

events - the former increases the plausibility of the message, and hence the difficulty of recognising it as phishing[9][10].

Campaigns can be carried out by the organisation itself - usually security staff, who may use an off-the shelf product they can configure, or by creating messages and the delivery mechanism from scratch. Alternatively, the organisation can commission an external service provider, who may then sent simulated phishing messages from within the organisation, or externally.

Phishing campaigns can be one-off or send repeated message over a period of time; they can target all employees or a subset, and the same messages may be sent to all, or messages may be targeted at subsets of employees. If a campaign involves multiple messages, the order may be random, or the campaign may start with the easiest or most difficult message, and increase or decrease in difficulty, respectively. Additionally, the level of difficulty of the next message sent may depend on whether the previous message was identified as a phishing one, or not.

Finally, there are several ways of how the organisation deals with the fact that such campaigns involve deception:

- The victim is informed once fallen for a simulated phish. In addition, the victim may or may not receive some information about phishing or explanations. If an explanation is provided, it can provide a form of "training by explaining", e.g. which signs in the specific messages could have been recognised as indicators of phishing.
- The organisation issues a general statement or message to staff that makes them believe that there has been a problem with the email service, but that does not disclose that they have fallen for a simulated phishing message.
- The victim does not realise that he/she has fallen for a phishing message, e.g. because he/she is redirected to a legitimate website.

In the last two cases, all staff or those who fell for the deception can be informed at a later date that they received phishing messages, and possibly be directed to further explanation or training.

In general, the organisation can inform staff that a campaign has taken place, with or without explanation. To do this by email is the most common form at the moment, but some organisations have provided explanation and results via departmental meetings or general staff meetings. This allows staff to engage and ask questions, and start a dialogue on how simulated phishing campaigns are perceived.

Phishing campaigns can be announced – beforehand – more or less prominently and in more or less detail. There may even be contexts where no information at all is given to the organisation's staff.

As for the survey of the current state (Objective 1), the evaluation of the teachable moment (Objective 2 b), or the evaluation of (newly developed) security awareness measures (Objective 3), different parameters are considered and collected either individually or in combination.

[9] Burns, Johnson, Caputo: Spear phishing in a barrel: Insights from a targeted phishing campaign. in Journal of Organizational Computing and Electronic Commerce 29(1):24-39

[10] Benenson, Gassmann, Landwirth: Unpacking Spear Phishing Susceptibility. Financial Cryptography Workshops 2017: 610-627

- The number of persons who per phishing message perform the relevant insecure action (e.g. click on the link/disclose sensitive data/open the attachment).
- The number of persons who report/delete an identified phishing message.
- The number of persons who, after having noticed it, report that they have fallen for a phishing message.
- The number of persons who are unsure and inquire about the received message.

Reporting can also be done at different levels of detail - for example, for all staff or for individual groups, or per phishing message or message type.

5 Problems with, and obstacles to, simulated phishing campaigns

First of all, we will consider reasons against organisations carrying out a simulated phishing campaign - irrespective of what the specific objectives of the campaign may be. The organisation may think about it as testing how resilient it is against such attacks, but staff may perceive it as being tested individually - and if they fall for a simulated phishing email, found wanting. The perception that they are being attacked by their own organisation while working to deliver its productivity goals can have a negative impact on staff trust in the organisation, and the security and error culture. This in turn can create a range of security problems. Furthermore, some aspects of simulated campaigns may not be compatible with national employment or data protection laws, or local agreements with labour organisations. All three aspects are discussed below.

5.1 Security Aspects

Simulated phishing messages try to reproduce messages sent by attackers, and do so with varying degrees of accuracy. Of the phishing messages mentioned in Section 2, those calling on their victims to make cash transfers or phone calls and those asking for security measures to be disabled or circumvented are more difficult to simulate. While the message itself is easy to send, it is difficult to verify whether an employee who received it thought it was genuine. Attackers sometimes send messages asking the recipient to deactivate or circumvent organisational security measures - simulated campaigns tend not to replicate this aspect because asking staff to do so poses a security risk. This has however a negative effect on how well the collected data reflects the organisation's actual vulnerability. An organisation needs to consider carefully what type of messages it considers a threat, and whether simulating those messages, and the actions some staff may take in response, create additional risk to the organisation.

Security problems caused by simulated phishing campaigns. Several security problems can arise while conducting a simulated phishing campaign. First of all, the infrastructure must be configured so that all simulated phishing messages reach the staff that is being targeted. In the case of emails, the simulated phishing messages must end up in staff's inboxes - if messages end up in junk or spam folders, staff will not

see them. This can happen especially if the messages are sent by an external provider of such campaigns.

There are a number of obvious questions: How can configurations be changed? Will the messages be tailored to each employee (considering his/her salutation) and will exactly these tailored messages then be whitelisted?[11] Or can only individual senders or individual names of attachments or domains/URLs be whitelisted?[12] In addition,the campaign provider usually has very limited knowledge of the infrastructure and can therefore only propose very general measures to make phishing campaigns possible. Thus, the risk of generally lowering the security level is correspondingly high. Is it technically possible to use whitelisting prior to the actual security audit of messages?[13] Should the security audit itself be adapted?[14] Organisations may end up creating configurations that reduce their level of protection in general. This is particularly risky because because genuine phishing messages will still be sent - and now are more likely to end up in staff inboxes.

Explanatory notes: (1) Whitelisting does not reflect reality - the following information would be missing: Are the phishing campaign emails really that ones that staff need to identify, because the security measures in place would not recognise and remove them? Or does the campaign mainly contain messages which, without whitelisting, would not reach the staff (and this would completely question the campaign's effectiveness)? (2) If the argument is that, for the simulated phishing campaign, one does not at all change the configuration, it must still be clear which of the campaign's messages actually reached which employee - otherwise, the effectiveness of the campaign again is questionable. If the argument is that almost all messages reach staff, one could argue that the organisation needs to improve its technical countermeasures to reduce its vulnerability. In addition, there is a productivity cost associated with staff with dealing with those messages, and running a campaign requires further resources, and is not without risks - so both in terms of security and economics, investing in better countermeasures makes more sense. (3) A change or reduction of the security measures for campaign purposes might violate organisations' security policies, and lead to significant problems during security audits.

An additional risk associated with a simulated phishing campaign is that attackers may use as a basis for a phishing attack. For example, the phisher can pretend to be the campaign organiser, and send a phishing email to all known email addresses. This email, for example, can contain a link or an attachment that is supposed to inform the recipient about his/her own performance in the campaign. Such attacks are possible, even if the staff have not been informed in detail in advance, since some staff end up alerting colleagues - either with the best of intentions, because they think it is an actual attack, or because they want to warn colleagues not to fall for the "same

[11] Wholesale general whitelisting means that phishers can take the same approach and can be sure that their phishing messages reach the recipient.

[12] This can still be used by phishers if this information is leaked.

[13] After the security audit, whitelisting is not helpful, because with an adequate security level of the security audit, most of the phishing campaign messages would be blocked and thus would present no risk to the organisation and its staff. This, again has a negative effect on how well the collected data reflects the organisation's actual vulnerability.

[14] If an external email service provider is used, this change may not be possible at all.

trick". If the campaign is run an external provider, the provider may name customer organisations in their advertising, and thus, staff may be forewarned, and attackers alerted to a possible opportunity. Organisations who use external providers should consider this carefully.

Explanatory note: Even if attackers have no concrete evidence that an organisation is running simulated phishing campaigns, they may try this because, as simulated phishing campaigns have become commonplace, staff are likely to have heard of them and thus believe that their organisation is, even if there has been no announcement.

Security problems associated with informing staff about simulated phishing campaigns Security problems can arise if staff cannot distinguish between simulated phishing messages, and real phishing messages that are sent by attackers, because staff might interact with a real phishing message (e.g. open the attachment, follow the link, etc.) - in most cases, they do it they feel invited to interact with the messages and learn more about them. Some staff may click on every link they see as a form of protest, because they feel it is unreasonable for the organisation to "trick" them in this way. No organisation can rule out that real phishing messages will reach their staff during the campaign - unless they stop delivering all external emails and only send simulated ones. If the organisation has promised there will be no negative consequences employees who fall for simulated phishing messages, they have to do the same for staff who fall for real ones. It is very difficult to communicate that, on the one hand, staff will not be punished for falling for simulated phishing, but at the same time, they are requested to be alert at all times, and try their very best to recognise and report phishing messages. The fact that it is not possible to distinguish between realistic, simulated and real phishing messages during a campaign means staff are faced with an impossible task - and we will return to the wider implications of this in the Human Aspects section below. If the organisation puts staff in that position, punishing staff who "fail" is from a legal and ethical point of view indefensible. Also, it reduces reporting, which plays an important role in reducing the damage of real phishing messages, because it enables the organisation to adjust its email filter and communication to staff.

A further problem is what happens after staff realise that they have fallen for a real phishing attack may not report it, because uncertainty or fear of negative consequences causes them to try to hide their failure and hope for the best. Others realise they have failed, but then assume it is part of the campaign, are be annoyed for a moment, but then don't report because they assume that their "failure" has already been logged.

Finally, whilst many staff are aware that distinguishing all simulated from real phishing messages is an impossible task, some are not. Such staff may be happy to have himself/herself identified a "simulated" phishing message, but may not report it because they have experienced a common side-effect of campaigns in the past: overwhelmed help-desks and IT security staff who respond with (with varying degrees of exasperation) "thank you, but it's a training message, no need to report but please don't tell anyone else". Negative experiences of reporting will lead staff to think twice before reporting again.

Explanatory notes: (1) Of course an organisation could point out that it is actually impossible to always distinguish between simulated and real phishing messages, and request staff to report all suspicious messages. But explaining why it is subjecting staff to training that has limited efficacy is challenging. (2) If the organisation is willing to this, the reporting and investigation processes must be clearly regulated, communicated, and integrated into everyday work. Staff must be given information on what to do, and to whom they should report if:

- they are unsure about a message
- they identify a phishing message, simulated or real
- they have fallen for a phishing message, simulated or real

Assuming the organisation has established such reporting and investigation protocol, and that in the run-up to the campaign, it made clear that distinguished between simulated and real phishing messages is not always possible, and thus reporting and investigation is mandatory.

Assuming a reporting and investigation protocol is in place, and staff understand and accept it, a phishing campaign will result in more reports and investigations. The situation can also become more complex because (a) it is now possible for staff to get feedback on which messages were simulated and real, (b) simulated and real phishing messages identified in time can be reported, and (c) staff who have fallen for a simulated or real phishing message can be told[15]. Rules for handling messages will be amended accordingly. Depending on the objective of the phishing campaign, documentation tasks will be added. To deal with this added load, the organisation has to either add staff to deal with these properly, or there will be longer waiting times. Longer waiting times will be frustrating for staff waiting to hear whether or not to proceed with a message, and increases the risk that they do interact with it. And again, delays in responses to reporting will reduce the likelihood of reporting in future. A reporting and investigation system that struggles to manage reports to real phishing messages will not be made better by the additional load created by simulated ones.

Explanatory notes: (1) We have witnessed several cases of simulated phishing campaigns in organisations being aborted or significantly scaled back because IT helpdesks and IT security staff were overwhelmed. In theory, the organisation should increase the number of staff for reporting and investigation during the campaign to be prepared for this. In practice, most organisations do not have suitable staff "on tap", and the administrative workload and financial implications of hiring temporary staff are off-putting - so some providers of simulated campaigns are reluctant to raise the issue, and play down the importance of reporting and investigation instead. (2) If there is any omission or ambiguity in the reporting and investigation protocol, there will be more enquiries because staff are unsure about how to deal with these messages and/or incidents.

Regarding Objective 2b (measuring if the combination of phishing campaign and subsequent security awareness measures has an effect). In order to measure the exact effect of

[15] Burns, Johnson, Caputo: Spear phishing in a barrel: Insights from a targeted phishing campaign. in Journal of organisational Computing and Electronic Commerce 29(1):24-39

the phishing campaign, no other security awareness measures should be provided to the staff at the same time. So, if the phishing campaign does not have the desired effect, the organisation has also lost the opportunity for other security awareness measures related to phishing.

In order to achieve the objective 2b, the campaign should last as long as possible, or even carry on permanently (to keep reminding staff of the risks of phishing). However, running campaigns on an ongoing basis would also require regularly reducing the level of protection (see above) - which most organisations will not want to do.

The level of difficulty of the simulated phishing messages can also have a negative impact on security. Messages that are easy to identify as phishing - e.g. with bad spelling and/or grammar, an unknown sender, or attachments with suspicious data types - lead staff to believe that they can detect phishing. This makes them more vulnerable to sophisticated attacks.

5.2 Legal Aspects

Organisations have to pay attention to the specific laws of the countries in which they operate and employ staff. In most cases (for Objectives 1 and 2b presented in Chapter 3), phishing campaigns will also measure work performance. This gives rise to legal questions in the context of employment protection and data protection. In Germany, for instance, the organisation's works council needs to be consulted on any measures to covertly assess staff. Data protection requirements may not allow identification of individual staff, in particular in European countries.

In addition to law pertaining to staff, legal limits may arise from trademark law. Messages pretending to be from other organisations, or clone phishing, may only look credible if they include the logos of the providers (e.g. of SAP or Paypal). It is only possible to do this without trademark infringement if campaigns are run purely in-house, on the organisation's own infrastructure, and do not express a business purpose of their own. However, all phishing websites created are then restricted for use within the organisation. It is also necessary to check whether the use of such logos conflicts with the organisation's code of conduct or other rules - imitating a trademark can be seen as a form of undermining that company's reputation, and make staff doubt it's trustworthiness in future. Intellectual property laws and copyright protection of such logos must be considered. All this needs to be clarified before phishing messages are sent. If clone phishing messages or messages from certain providers cannot be used, this limits the significance of the phishing campaign, as it is not possible to send the kind of phishing messages that are common otherwise.

5.3 Human Aspects

In most organisations, security is not a particularly popular topic with staff. Most organisations have many security policies, which staff may have heard or read about, but often find impossible to understand or follow. Staff feel justified in ignoring security policies and training that they find impossible to follow and/or that noticeably reduce their productivity. With simulated phishing campaigns, staff are taught to check many aspects of each message they receive - one campaign commissioned by a UK

bank literally tells people to ”take 5” - minutes - before acting on a message. Since most staff in modern organisations receive dozens or hundreds of emails per day, the productivity reduction that would result from following this advice would be enormous. Many organisations have security policies create impossible tasks, with serious side effects: unsurprisingly, it creates resentment, and creates the perception that IT security is ”impossible” and best ignored. Yet staff do not happily breach the rules - they do worry about enabling a breach and being blamed for it, after having failed at the impossible task. Simulated phishing campaigns contribute further to this unhappy state of affairs.

In a phishing campaign, the organisation ”attacks” its own staff. In particular, if the campaign is conducted internally, one group of staff (usually from the IT or security department) attacks all others - though they may not attack the organisation's management. Depending on the design of the phishing messages, all staff attack each other to a certain extent: since it is impossible for staff to distinguish whether the received message is part of the simulated phishing campaign, a real phishing attack, or whether the colleague wrote the message based on a message he/she received within the campaign. As soon as word of the latter possibility has spread among the staff, distrust among staff increases, and already difficult relationships will deteriorate further. This, in turn, also leads to reduced productivity, and in the worst case, mediation talks will become necessary.

If a campaign is announced, the sender will receive questions. If the sender is a member of the IT or security department, and does not have a good relationship, they might misinterpret those questions, which again would lead to new conflicts. If the campaign has not been announced, and a simulated phishing message is identified as a phishing message, staff members may feel tricked, in which case resentment and potentially conflicts may follow. Phishing campaigns, in which messages from other staff - e.g. the CEO or Head of Human Resources - are simulated, it may affect the perception of those staff members, trust in the organisation. Staff who fall for (poorly designed) phishing messages might be perceived as stupid or careless, and treated with disrespect. This also has a negative effect on the culture of the organisation.

Launching a phishing campaign without first instructing staff (i.e., explaining how to identify phishing messages, where to report if they are unsure, how to deal with phishing messages they have identified, and where to report if they fell for such messages) is simply unfair. Simulated phishing campaigns are not likely to increase staff' trust in the organisation's management - particularly if the campaign is not widely announced. Simulated campaigns are supposed to improve the security awareness and skills of staff, but those who fall for a simulated message experience failure. Self-efficacy has been shown to be the key factor in changing security behaviour: staff who are confident in their ability to perform a new security task are significantly more likely to change behaviour than those who are not. Making staff experience failure a security task can lead them to conclude they cannot reliably detect messages anyway, and thus resort to reporting any message they are not absolutely sure of.

If the campaign has not been announced officially, but rumours are circulating through the organisation, further problems arise if the reporting and investigation processes have not been clearly communicated in advance of the campaign: Staff are

unsure of how to proceed if they have identified or fallen for a phishing message. Does the same reporting process apply to simulated phishing? Why do I get this message and others do not? Who knows now that I have made a mistake? What is the consequence of this? Often, staff feel unsettled and controlled by phishing campaigns. Both will have a negative impact on the organisation's reporting culture. Another problem of not announcing the campaigns officially and in detail is that false information then spreads quickly and is difficult to correct. Simulated phishing messages have to be treated in the same way in the reporting and inquiry process. This not only requires a great deal of effort and resources, but deceiving staff. Deception does not create trust in the reporting and investigation system. This can even have a negative effect on the level of protection if the consequence is that people generally do not want to ask questions or do not want to report anything because they do not want this kind of behaviour. Accordingly, phishing campaigns also have a negative influence on the organisations' error culture.

If the phishing campaign is widely announced, staff will look critically at many more messages, and may try to verify the sender, e.g. by phone. This again reduces productivity, and being asked several times an hour whether I really sent that messages will not improve working relationships. organisations cannot expect work to continue as normal during a simulated campaign - staff must be given time to deal with the extra work of scrutinising messages and reporting them. If this is not the case, it increases pressure and has a negative effect on the perception of the organisation's leadership.

Further productivity losses can result staff becoming overly cautious, and treating legitimate messages as phishing messages. This means invoices are not paid, job applications not considered, and queries by customers or suppliers ignored. Message from any 3rd party - e.g. an external travel or survey company - now have a hard time getting responses from staff in many organisations. A responsive reporting and investigation process that can quickly provide responses can help, but it requires significant resource.

The level of difficulty of the simulated phishing messages can also have a negative influence on staff' mood. If they are too easy, the impression can quickly arise that the management thinks that staff cannot be identify obvious phishing attempts. If an employee receives a second phishing message of the same type (after one has fallen for the first one of that type), she gets the impression that the management suspects that she still does not understand. Again, this does not have a positive effect on staff' trust in the management.

Regarding Objective 2b (proof was given of the fact that the combination of phishing campaign and subsequent security awareness measures has an effect). From the point of view of the phishing campaign alone, this would mean that staff would only have access to the information from the security awareness measure if they had made a mistake. This easily leads to irritation and uncertainty. It leads to the situation that if you want to know more (because you think you do not know enough) you have to interact with phishing messages, which, as one knows, should actually not be done. This has a negative effect on the self-efficacy of staff and reduces the level of protection.

Regarding Objective 1 (assessing the current situation and then motivate for a subsequent security awareness measure) and Objective 2 b). These cases entail another problem: Nobody likes being confronted with his/her own weaknesses. Being told what you do wrong makes you feel bad. However, that's exactly what may happen if phishing campaigns are designed in such a way that sooner or later one finds out that one has fallen for a phishing message. Very well-designed phishing messages are very likely to have many victims, and they initially go through a negative experience. It is questionable whether and how the willingness to learn how to identify phishing messages in the future will increase as a result of a negative experience; it has been shown several times that a lack of self-efficacy has a negative influence on security behaviour[16]. In the case of Objective 2b, it is particularly important to find out whether the staff who became victims were so shocked and surprised by their own failure that they closed the document or exited the website quickly, so that no one else will notice that they have fallen for a phishing message. Accordingly, the victims would not notice that information on how to recognise phishing messages in the future is provided in the message.

6 What do the numbers collected during the simulated phishing campaign tell us?

Data regarding the people having fallen or having (not) reported are collected in case of Objectives 1, 2 b), and 3. To achieve theses objectives, different data types - individually or in combination - can be considered, i.e.:

1. the number of staff who perform the corresponding unwanted action per phishing message (e.g. click on the link/open the attachment); this needs to be defined more precisely (e.g. clicking on a link only, or entering access data or other sensitive data[17],
2. the number of people who report a detected phishing message,
3. the number of people who report that they have fallen for a phishing message after they have discovered the deception,
4. the number of people who respond with inquiries about messages.

The question is what can cause high or low numbers of people who fall for a simulated phishing attack.

The (external) validity of the results is strongly influenced by several facts. First of all, the external validity depends on the amount of information about the phishing campaign distributed to the staff. A (large) part of the staff will be more sceptical about the relevant messages than usual, will ask or inform colleagues, or generally talk about it having discovered a phishing message. Others are so against "attacking" staff that they intentionally interact with every phishing message.

[16] https://www.enisa.europa.eu/publications/cybersecurity-culture-guidelines-behavioural-aspects-of-cybersecurity

[17] The latter quickly becomes another security problem. Because these sensitive data must not be transferred.

All this is especially true when the time frames for campaigns are short. At the same time, short campaigns only increase the security risk for a short period of time. All these influencing factors should at least be considered as a limitation of the validity[18].

It is also assumed that there is a well-established reporting and inquiry process even before the phishing campaign starts. It would also be necessary for the process to provide for the reporting of detected phishing messages, and not for how many have been deleted. Otherwise, phishing messages cannot be distinguished from other spam messages or other messages that are deleted. Part of the reporting and inquiry process must also include the reporting of phishing messages that are already known to have been reported by other staff (e.g. a colleague working in the same office).

Explanatory note: It is strongly recommended to first critically analyse one's own reporting and inquiry processes before considering a simulated phishing campaign.

In case you are wondering why it is not sufficient to record the number of staff who perform the relevant unwanted action per phishing message (e.g. click on the link/open the attachment), please consider the following explanations.

Non-interaction can have many reasons and therefore cannot be interpreted as a clear indicator that the message was identified as a phishing message: The message was not seen at all because the person concerned was on vacation or sick, had no time or was not relevant, does not have a corresponding account, or because a colleague has already drawn attention to the phishing message. The latter does not mean that the respective colleagues would have automatically identified the message as a phishing message. It is not possible either to tell staff not to inform other staff, because this is exactly what you want in the case of real phishing messages: Staff are supposed to react and help others to protect themselves and the organisation.

Ultimately, the false positives would also have to be counted, i.e. messages that were legitimate but were reported as a phishing attack and were therefore not processed for the time being. To put it bluntly: A phishing campaign that has the consequence that all phishing messages are reliably detected, but that also has the consequence that every second legitimate message is deleted because it is considered to be a phishing message, is not effective either.

The external validity of the results of a phishing campaign also depends on the simulated phishing messages. The rule of thumb is: The easier they are to detect, the "better" the results. Simulated phishing messages that are extremely difficult to identify would hardly be recognised as such by anyone. Actually, the simulated phishing messages would have to represent those from real attacks (and thus a multitude of different ones), but for this purpose, messages from staff and external providers would also have to be used, which would have a number of disadvantages (including the relationship of trust between colleagues and the fact that one might have to check trademark law aspects). All in all, the external validity should always be seen in relation to the simulated phishing messages as well as in relation to changes in the infrastructure.

[18] This is particularly true in the case of Objective 3, and any other evaluation would also have limitations. Here, it may make sense to use different study forms for the evaluation.

The results' validity of the collected data also depends on whether or to what extent other influencing factors, e.g. media reports, can be controlled during the period of the survey.

If Objective 2(b) is pursued, the following should be considered in addition:

Additional data would need to be collected. For the evaluation, it would be necessary to know whether and how long the subsequent security awareness measures have been dealt with. For this purpose, the data could only be collected in a pseudonymised but not in an anonymised way. This must be checked for admissibility under data protection law.

7 Conclusion

The external validity of results for simulated phishing campaigns in general, and especially for some particular forms, is a matter of debate. Security experts and service providers selling the services equate a reduction in click rates with reduction in vulnerability, while human factors experts point to the futility of training staff on what is essentially an impossible task, and economists - and some organisational leaders - count the mounting productivity losses caused by this countermeasure.

Our analysis has shown that creating a simulated phishing campaign that minimises additional security problems and is legally compliant is extremely difficult and costly. Even if an organisation is willing to invest this much, the combined negative impact of simulated phishing campaigns on the self-efficacy of individual staff, and the reduction of inter-personal trust and trust in the organisation, and the reduced productivity of all staff involved are enormous. The cost and negative side-effects clearly outweigh the low external validity of a such a campaign, and the limited reduction in vulnerability that results. We therefore recommend that organisations invest time and money in (1) an improvement of technical measures. In addition, (2) appropriate awareness measures should make staff aware of the type of phishing messages they can reach despite all technical measures and of how they can identify them. Finally, (3) the reporting and inquiry process should be improved. As a result, the effort for each individual employee is comparably low and can be implemented. The level of protection increases without negative effects on trust relationships and self-efficacy.