On the Need for New Anti-phishing Measures Against Spear Phishing Attacks

Luca Allodi, Tzouliano Chotza, Ekaterina Panina, Nicola Zannone
Eindhoven University of Technology
l.allodi@tue.nl, t.chotza@student.tue.nl, {e.panina,n.zannone}@tue.nl

***Abstract*—We provide an extensive analysis of the (unique) characteristics of phishing and spear-phishing attacks. We argue that spear-phishing attacks cannot be well-captured by current countermeasures, and identify ways forward. We exemplify this by analyzing an advanced spear-phishing campaign targeting white-collar workers in 32 countries worldwide.**

***Index Terms*—Social engineering, phishing, spear-phishing, countermeasure effectiveness**

I. Introduction

The advanced technical defensive capabilities of modern systems are continuously increasing the burden of malware and exploit engineering, who have to circumvent more and more sophisticated mitigation techniques such as (Kernel) Address Space Layout Randomization (K-ASLR), Windows Defender Exploit Guard, and many others. The increasing difficulty associated with exploit engineering manifests, for example, in the relatively low number of marketed exploits and their time of appearance [1]. For this reason, social engineering is becoming a predominant attack vector – targeting the human rather than the system – and therefore a critical issue for organizations; recent figures suggest as much as 76% of organizations have experienced spear-phishing attacks in recent years [2].

Differently from phishing, spear-phishing is a highly targeted, context-specific attack that is directed at specific groups of individuals or organizations [3]. The contextual dimension of spear-phishing attacks is relatively unexplored in the literature, and is currently unclear which factors make spear-phishing attacks successful [4]. Spear-phishing attacks are generally characterized by a multi-stage process whereby the attacker collects information on a specific target or group of targets. In contrast to one-size-fits-all phishing, spear-phishing attacks are engineered to fit specific victim profiles; for example, information on previous actions or (likely) decisions undertaken by the victim can be weaponized by the attacker to forge effective social engineering artifacts. However, in real scenarios a clear-cut distinction between phishing and spear-phishing is oftentimes difficult to make; for example, unsophisticated attackers after a specific group of targets may still launch 'generalistic' phishing attacks, despite having access to highly targeted information on the victims [4]. In this paper, we consider spear-phishing an attack that actively employs collected information on its targets to forge the attack artifacts.

On the defender's side, current best practices and technical solutions for phishing prevention, detection, and mitigation do not (and in some cases cannot) account for human-related characteristics, which are typically exploited in spear-phishing attacks. The lack of specific awareness on the peculiarities of spear-phishing attacks is limiting advancements in this field and must be addressed before new countermeasures against spear-phishing can be developed and tested.

In this article we fill a gap in recent literature by highlight and systematically analyze the spear-phishing phenomenon. Our contribution is threefold:

- we perform an extensive literature review dissecting and differentiating the attack process and characteristics of spear-phishing attacks from regular phishing;
- we identify a foundational gap between existing countermeasures tackling phishing, and the features and *modus operandi* of spear-phishing attacks that must be addressed by future research;
- to exemplify spear-phishing attacks' unique features, we analyze a real, advanced campaign targeted at white collar workers across 32 countries worldwide that happened at the end of 2017/beginning of 2018.

The paper is organized as follows. The next section provides background knowledge about social engineering attacks and countermeasures against such attacks. Section III introduces our methodology and Section IV presents an analysis of phishing and spear-phishing attacks. Section V analyzes a real spear phishing attack. Section VI concludes with an analysis of the relation between anti-phishing measures currently employed by organizations and (spear-)phishing features, and discusses lessons learned.

II. Social engineering techniques and countermeasures

This section presents an overview of social engineering attacks and the countermeasures typically employed against such attacks.

A. Social engineering techniques

Social engineering attacks can take diverse forms (phishing, physical interactions, phone calls, etc.), and usually aim at stealing sensitive or confidential information (e.g., banking information, passwords, and credit card details) by mimicking electronic communication from a trusted source. A bird's eye view of the overall attack process is given in Figure 1. A social engineering attack can happen in multiple stages, composed of several phases each. Phases executed by the attacker are

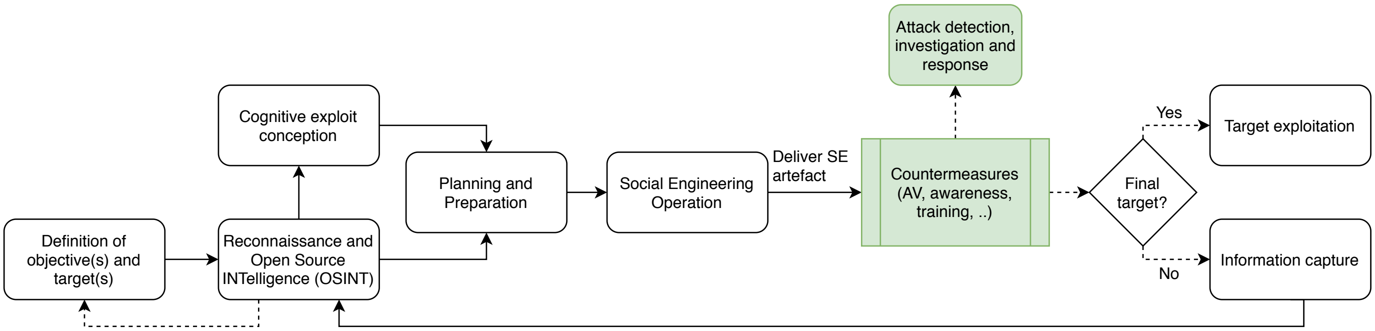
Attack detection, investigation and response
Cognitive exploit conception
Planning and Preparation
Social Engineering Operation
Deliver SE artefact
Countermeasures (AV, awareness, training, ..)
Yes
Target exploitation
Final target?
No
Information capture
Definition of objective(s) and target(s)
Reconnaissance and Open Source INTelligence (OSINT)

Fig. 1. Schema of social engineering attack phases and role of countermeasures.

depicted with black-lined boxes. Activities and countermeasures adopted to counter the attack are represented in green.

1) Attack stages: During the first phase, the attacker establishes specific attack objectives (e.g., stealing credentials, obtaining sensitive information) and identifies potential target(s) of interest. Once the potential target(s) are identified, the attacker may collect additional contextual information, generally from open-source and *external* resources. Based on the intel gathered during the reconnaissance and OSINT phase, the attacker may refine the set of targets for the planned attack. For example, the attacker may add new discovered targets, or find intermediate targets that can be exploited to reach the final objective (e.g., moving through an organization's hierarchy).

Once the objectives and targets are defined, the attacker gatheres intel on the targets and uses it to devise the cognitive exploits and to plan and prepare the attack. In this phase, the attacker develops the attack strategy and the artifacts needed to implement it. These have to be suitable to the target environment and w.r.t. to the final objective of the attack. Some artifacts may also be developed to support the context or the pretext of the attack. For example, online resources can be crafted to support the attacker's claims, including websites, Wikipedia entries, or forged identities. The artifacts may be crafted specifically for the victim system, accounting for system configuration (e.g., to maximize malware success probability) or for the specific environment in which the victim normally operates (e.g., professionally, type of language used in official communications).

Finally, the attack artifact(s) are delivered to the devised recipient(s) in the social engineering operation phase. If the recipient is the final target, the attacker waits for the attack payload to be executed on the victim side. Otherwise, the attack can be cycled through in as many subsequent stages as needed, during which the attacker collects additional, more targeted information about the victim, and escalates from there until it reaches their final objective.

2) Cognitive exploits: Crafting believable social engineering artifacts requires attackers to take into account several aspects related to human cognition and psychology [5]. These aspects are often related to specific characteristics of the (human) victim of the attack, for example determined by their social context, or professional affiliations. The techniques adopted to forge the human-interaction artifacts are a key element for the success of the attack [6]. For example, attackers can exploit *brand theft* to craft social engineering attacks by cloning the logo and style used by well-known and trusted public and private organizations, and spoofing their identity in electronic, printed, or phone communication. A recent example involved numerous scams against foreign immigrants in the Netherlands where scammers performed phone calls pretending to be originating from the IND, the Dutch immigration service [7]. The *language* used to craft the attack is also a critical factor for its success. Interactions happening using the first language of the victim and tailored to her organization are potentially more persuasive due to the sense of familiarity that is established [8].

Individuals can be manipulated by exploiting their habits, motives, and cognitive biases, such as curiosity, anger, fear, patriotism, friendship, altruism, vanity, community effects and sense of duty [5], [9]. Cialdini [5] has identified six fundamental principles of persuasion that can be exploited to influence individuals, namely *liking*, *reciprocity*, *social proof*, *consistency*, *authority* and *scarcity*. The effectiveness of these cognitive vulnerabilities in the context of social engineering attacks has been largely studied in the literature. Some vulnerabilities like *liking* and *scarcity* have been proven to be effective in increasing the attack's success probability. Other vulnerabilities like *social proof* and *authority* have been shown to be effective only in some cases. For instance, while most people are inclined to follow recommendations and suggestions from authorities, authority can be perceived negatively by individuals when threatening their freedom of choice. Also, *reciprocity* has been shown to not have a considerable influence on individuals [9]. This is mainly because this cognitive vulnerability oftentimes requires constructing a fictitious situation in which the victim feels the obligation to reply, which may be very challenging to achieve.

Personal characteristics of the victim have also an effect on attack success rates. Several studies have analyzed individuals' susceptibility to social engineering based on *gender* and *age* of the victims [2]. The familiarity with the communication within the organization and the *awareness* (e.g., through experience or training) has been shown to have a positive effect, however with varying degrees of success [6]. These contrasting results can be traced back to other victims' characteristics related, for instance, to their *location* and *culture*.

B. Countermeasures

A number of countermeasures against social engineering attacks have been proposed in both academia and industry. Following the classification defined by the NIST Cybersecurity Framework, we group them in three main classes: *protection*, *detection* and *response*.

1) Protection: Organizations usually offer their employees security-awareness *training*. Such training can have different scopes and rely on different levels of preexisting knowledge. Some training aims to increase awareness on the threats posed by social engineering attacks, and the risks of sharing personal information (e.g. on social media), for both themselves and the organization.

Access control is employed to restrict employees' access to only the resources they need to perform their work-related activities. This may mitigate attack opportunities by minimizing the attack surface to which each employee exposes the organization. Organizations also employ *information protection procedures* to impose restrictions on which information employees, contractors and suppliers can disclose about the organization. *Protective technologies* (e.g., blacklisting, application firewalls, URL sanitization). are also used by organizations to block attempts to reach phishing and, in general, malicious websites.

2) Detection: Organizations typically have a detection process in place to detect social engineering attacks, including phishing and phone scams. This process provides continuous monitoring capabilities to detect anomalous behaviors within the organization network. *Network monitoring and filtering* tools employed in this process aim to detect attacks artifacts, identify spoofing attempts, and deceiving websites.

3) Response: The ability to timely respond to social engineering attacks, most prominently phishing, is crucial for organizations. *Incident reporting* provides a means to promptly notify (e.g., through advisories) interested parties of incoming attacks in order to reduce their effect. Large organizations such as banks or financial operators can also employ response teams whose *reason-d'être* is to analyze and *mitigate* incoming attacks such as fraud, scams, and phishing. These teams usually operate in security operation centers and rely on detection technologies outlined above to prioritize their work [10] in order to react to the attack (e.g., by means of takedown actions, or contacting affected customers).

III. Evaluation criteria

In this section, we provide a breakdown of the procedures and criteria employed in social engineering attacks to establish a baseline to: 1) compare phishing and spear-phishing attacks over a defined set of dimensions characterizing their core aspects; 2) match existing countermeasures against different attack features and characteristics. We have identified three core dimensions characterizing social engineering attacks: *deployment*, *cognitive features*, *victim demographics*.

a) Deployment: We identify three main criteria:

- *Attack scope*: Social engineering attacks at large can be aimed at large populations of Internet users or at pre-selected specific users or groups of users. In the first case, victims' only common characteristic is that they belong to the same 'community' or pool of Internet users (e.g., users of a social networking platform, or clients of a bank); in the second case, attackers pre-filter users by selecting specific targets with certain desirable profile characteristics.
- *Sophistication level*: Social engineering attacks can either be executed in a one-shot interaction between the attacker and the victim, or over multiple interactions. In one-shot interaction attacks, the attacker has only one opportunity to lure the victim into compliance (e.g., to click on a link). Conversely, multi-stage attacks break the process over multiple interactions, at the price of an increased attack cost for the attacker.
- *Artifacts*: Social engineering attacks can rely on a number of artifacts for the delivery of the attack, and the form that the payload (delivering the final impact on the user) takes.

b) Cognitive features: These features characterize how attackers increase their chances of persuading victims into falling for the attack. The literature identifies six 'cognitive vulnerabilities' that formalize *how* these features can be characterized:

- *Liking*: by providing clues for which the victim will pose higher trust in the attacker, victims can feel more inclined to comply to requests.
- *Reciprocity*: by providing unrequested favours or promises to the victim, the attacker can trigger a response by encouraging the victim in *returning* the favour or the kind act by complying to the request.
- *Consistency*: victims are more inclined in sticking to previous decisions to which they already committed; attackers can refer to past (fictitious or not) interactions to increase their chances of a positive response.
- *Social proof*: under uncertainty, victims are more inclined to *copy* the behaviour of their peers. The attacker can convince the victim that similar users have already taken a certain decision to convince the victim that that is indeed the right thing to do (e.g., renewing a password).
- *Authority*: attackers can exercise authority over victims to impose a fear of punishment wherever the victim decides to *not* comply with the requested action.
- *Scarcity*: victims are more inclined to act irrationally (in the economic sense) when they are given only little time to take a decision; in this scenario, victims tend to not correctly weight the information they receive in the fear of *missing out*, or of suffering from high opportunity costs.

c) Victim demographics: The type of users that are targeted plays a role in how a phishing campaign can be expected to be successful. This dimension is different from *attack scope* as users with similar demographics can belong to the same large pool of users without any specific pre-filtering on the attacker side. These variables include *personal* characteristics (e.g., *age*, *gender*, *culture*, and *location*) and *professional* characteristics (e.g., *role*, *awareness*, *years of service*).

IV. Breakdown of phishing and spear-phishing

Even though phishing and spear phishing campaigns have similarities, spear phishing campaigns are typically more sophisticated.

A. Phishing

1) Characteristics of deployment: Phishing attacks are usually of a 'hit-or-miss' nature whereby only one attempt by the attacker is made (i.e., a single-stage of the attack process in Figure 1). This is consistent with an attack model whereby the attacker is not focused on a specific victim, but relies on 'large numbers' of potential victims to collect a sizable return (in terms of stolen credentials, infections, banking details, or else) from the attack. These attacks generally involve the development of two artifacts: the phishing vector (an email, voice call script, or other form of social interaction), and a phishing payload (a malicious attachment, or a phishing domain). The development of the phishing vector can have varying degrees of sophistication [11], from simple 'hooks' to more sophisticated combinations of cognitive attacks [9].

2) Cognitive features: The cognitive features used in phishing may vary depending on the domain from which users are pooled from, and their effectiveness may vary accordingly. Common choices across domains are *Scarcity*, *Social Proof*, *Liking*, and *Authority* [6]. Other cognitive vulnerabilities such as *Consistency* and *Reciprocity* are less applicable in the context of phishing attacks. Introducing *fictitious prior shared experiences* is especially hard in this context as fabricating an experience between attacker and victim may increase the chances of creating discrepancies between the 'fictitious experience' and the 'actual experience' the victim can relate to. This is particularly difficult in generalist attacks or attacks targeting a broad pool of users, where a credible personal dimension cannot be easily integrated in the attack.

3) Victim demographics: Victims of phishing attacks are often 'pooled' from leaked or stolen user data from specific domains (e.g., from the bank sectors, or clients of a specific service), and can be bought or accessed through online criminal resources [1]. Victims of phishing attacks have only few circumstantial (as opposed to profile-specific) traits in common. This, however, provides a credible venue for attackers to decide on which attack artifact to develop, and how (a banking email or a commercial offer). In more generalist attacks, phishers do not target a specific pool of users and cannot rely to well-defined scenarios for their attack. For this reason, attacks are generally not tailored around a specific victim profile, but rather targeted to a broader and variegate audience.

B. Spear-phishing

1) Characteristics of deployment: Spear-phishing campaigns are created specifically for an individual, a small group of individuals or an organization. For this reason they are characterized by a more complex attack process than normal phishing, whereby iterative information collection (*reconnaissance*) and attack engineering is employed to maximize the chances of a successful payload delivery. In particular, the attack phases in Figure 1 are executed as many times as needed in order to obtain the required goals such as information, access and/or privileges.

2) Cognitive features: Due to the different attack dynamics, in spear-phishing attacks the attacker is more flexible in choosing and distributing cognitive attacks over the attack process. Importantly, due to the multi-stage nature of the attack, the attacker has the capability of constructing previous interactions with the victim, which can constitute a basis for further attacks; for example, the attacker may leverage previous decisions taken by the victim during that interaction to increase the likelihood of compliance to a follow-up request (principle of *Consistency*). This spreading of cognitive vulnerabilities across several interactions produces different and more variegate attack profiles w.r.t. what one can expect from an 'ordinary' phishing attack.

3) Victim demographics: The target of a spear-phishing campaign is generally very well identified by the attacker *before* starting the attack. The information collection phase (ref. Figure 1) allows the attacker to fetch intelligence regarding the profile, habits, and social surroundings of the victim. Importantly, by means of this evaluation, the attacker can also infer the *circle of trust* around the victim, and identify possible weak points to enter it. Across the whole chain, attacks can be tailored specifically for the profile of that stage's victim, exploiting personal weaknesses or professional ambitions, for which target demographics are a good predictor.

V. Breakdown of a real spear phishing attack

In this section we analyze a highly-targeted spear-phishing campaign against white collar workers on LinkedIn. This case study emerged from a contingency involving one of the authors of this paper (below 'the applicant'). The reported names of the involved (real) companies are censored.

Among other open positions, the applicant applied to an open job posting for a *Project Manager* position at `Eliora Construction`, which later revealed to be a fictitious company created specifically for the perpetration of the spear-phishing campaign described here. `Eliora Construction` poses itself as a company operating in the constructions sectors (asphalt and roads) for more than 90 years, with a significant international profile and presence in 17 countries worldwide. `Eliora Construction` lists its offices to be located in Charlotte, in North Carolina, US. The opened job position aims at the medium-high sector of the job market, calling for mid-seniority (5-10 years of relevant professional experience) personnel with previous managerial experience. An identical position and job description was provided for 32 locations worldwide, spanning from European countries to Asia, Australia, and African countries, whereas no position was opened in the US. The application was done on the 23rd of Nov. 2017.

1) Characteristics of deployment:

a) Decoy website and affiliated companies: On `Eliora Construction`'s website the company mentions to have a parent company, `ParentC`, a leader in the sector. A visit to `ParentC`'s website reveals the list of worldwide partners of the company by region. Interestingly, `ParentC` only reports

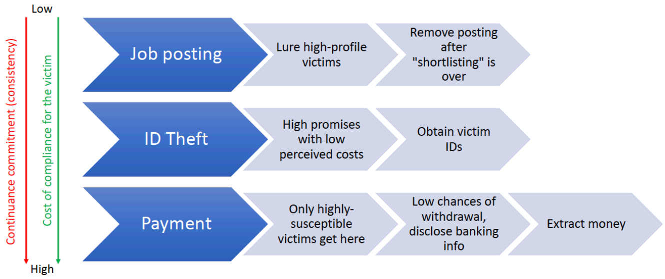
Low
Continuance commitment (consistency)
Cost of compliance for the victim
High
Job posting
Lure high-profile victims
Remove posting after "shortlisting" is over
ID Theft
High promises with low perceived costs
Obtain victim IDs
Payment
Only highly-susceptible victims get here
Low chances of withdrawal, disclose banking info
Extract money

Fig. 2. Breakdown of attack steps in the analyzed spear-phishing campaign

two affiliate companies in the US: `SpoofedC`, in North Carolina, and a second company in Florida. Visiting their respective websites, it is apparent that `Eliora Construction`'s webpage is a perfect clone of `SpoofedC`'s. Content-wise, the two websites appear to be identical, with the exception of the reported company's name throughout the text, and less recently updated 'news' entries on `Eliora Construction`'s website. Further inspection of the page's source code reveals that `Eliora Construction`'s website has been cloned with `HTTrack`, a popular open source tool to dump entire websites by traversing hyperlinks and `src` pointers in a webpage. A `whois` lookup for the website also reveals that `Eliora Construction`'s domain was registered just a few days ahead of the campaign. A final check on the North Carolina Secretary of State's service website reveals that no company is registered under the name of '`Eliora Construction`'.

b) Attack phases: The attack was distributed over three phases: recruitment, ID theft, and Payment. Figure 2 provides a breakdown of the three phases of the attack.

(1) Recruitment through job posting. During the recruitment phase, victims were lured into sending in their curriculum and information thought the LinkedIn procedure. This phase constitutes a first level of engagement with the potential victim and is within the expectations a user of that social platform has under these circumstances. A week later an invitation letter for an interview was delivered to the applicant. The letter was sent and signed by the Human Resource Manager of `Eliora Construction`:

> Dear Applicant, I write to inform you that your resume has been properly reviewed and screened by our recruiting board and you have been found eligible for this vacant position. Be informed that you have been shortlisted for an interview scheduled for Friday, 12th of January 2018 at ELIORA CONSTRUCTION COMPANY, 1055 Metropolitan Avenue Charlotte, North Carolina, 28204, United States of America.

The letter continued by specifying the job reference number a thorough description of the job position that was included to underline the "*important roles/responsibilities expected of a Project Manager by ELIORA CONSTRUCTION COMPANY*", and specifying:

> [..] our primary reason for requesting for your physical presence is to have our chief project manager have a one on one interview with you and ensure you possess the aforementioned qualities and also have you familiarize yourself with the company structure as well as a recap on past and upcoming project.

(2) ID theft and (3) pre-payment. The payload of this attack is distributed over two subsequent phases, the second of which can only be achieved after a full commitment (on the side of the victim) for the former. This is accomplished by the attackers by providing the following information to the victims:

> Please note that **our official travelling consultant shall handle your travel needs** which will include flight tickets, hotel reservations, **visa procurement** and transfers within the United States. More so, **you will be responsible for all your travel expenses** made through our affiliated travel agency. These expenses shall then be refunded to you by `Eliora Construction` on arrival at the interview venue.

The invitation letter outlines the procedure to follow for the job interview. Importantly, the instructions explicitly remark that *first* the victim has to send in their VISA/travel information, including a copy of the passport and a photograph, and only at a subsequent moment the payment details will be disclosed to them.

2) Cognitive features: The multi-phase stages of this attack allowed attackers to build on top of previous interactions with the victims. For each step of the attack (ref. Figure 2), the attackers ask increasingly higher levels of compliance relying on previous commitments the victim has, at that stage of the attack, already embraced: during the first stage, attackers collected victim information and professional profiles through the usual LinkedIn procedure. This constitutes a first, low level of compliance to which the victim, willingly, agrees. Leveraging from the compliance obtained during the first stage, the attackers can then step further to achieve the first objective of the attack, namely the identities and passports of high-profile white collar workers. This second stage comes in, again, naturally following the invitation letter, and exploits the *Social proof* and *Authority* principles in that the victim already recognizes that this is the normal behaviour requested to any applicant in this situation. Finally, once the victim *already* complied to the disclosure of his or her identity to the attackers, the payment is requested. Table I provides an excerpt of cognitive attacks embedded in the communications between attackers and victims.

3) Victim demographics: We exploited a LinkedIn Premium subscription for an a-posteriori analysis of victim demographics. In particular, we collected anonymous, aggregate information regarding the other applicants, their professional and academic profiles, and country of origin. The collected data show that the attack was targeted specifically at white-collar workers from specific countries with a mid-high professional seniority levels and high academic profiles. Figure 3 reports an excerpt of the countries from the Eurasian and North African regions targeted in the campaign. Targeted countries outside these regions are: `Colombia, Singapore, Malaysia,`

TABLE I
EXCERPT OF COGNITIVE FEATURES EMPLOYED IN THE REPORTED SPEAR-PHISHING CAMPAIGN

Reciprocation	Consistency	Likeability	Scarcity
We demand people who will thrive in a culture that encourages and rewards innovation, collaboration and the ability to learn from success as well as failure.”	Job Locations: As advertised on LinkedIn (Further information will be issued after the interview).	Has a high reputation for innovation in civil engineering. founded in the 1920s, ..is one of North Carolina’s leading construction companys”	Interviews are also designed to ascertain claims of working experience. Should any claim be found wanting the affected expatriate may be deported
Expatriates are entitled to a one (1) month paid home leave every twelve (12) months. [..] All expatriates are entitled to free two-way tickets to cover the span of their home leave.	Basic Salary: Between $105,000.00$ USD and $160,000.00$ USD per annum (this is will be disclosed after interview).	Today, we have interest in design & project management ventures in more than 11 countries and to employ approximately 2000 staffs (not including contractors).	Please note that our official travelling consultant shall handle your travel needs [..] you will be responsible for all your travel expenses made through our affiliated travel agency.
All expatriates are entitled to a free and mandatory safety courses on Job Locations to be delivered by qualified safety and environment experts.	Our company’s accountant will furnish you with our banking details for making a wire transfer of your booking cost as soon as your documents have been received.		Date of Interview: Friday, 12th of January 2018

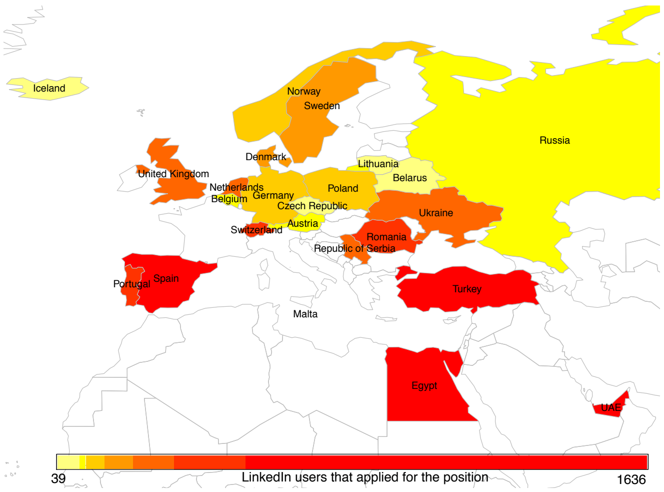

Fig. 3. Excerpt of countries targeted in the campaign. We report only the Eurasian continent and North Africa for legibility.

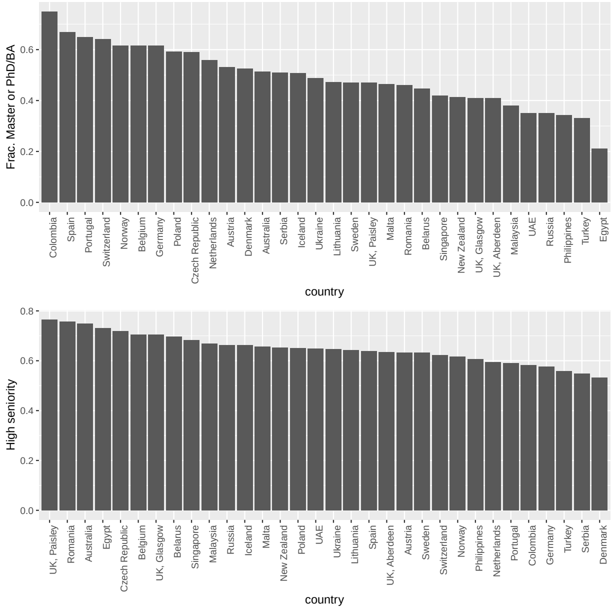

Fig. 4. Application response rates per country by (top) academic qualification and (bottom) professional seniority.

`Philippines`, `Australia`, `New Zealand`. This selection of targets gives the opportunity for the execution of the *ID theft* phase of the attack, by providing the pretext to ask for victims’ IDs for the preparation of the travel documentation. Figure 4 provides an overview of the academic and professional seniority of the victims targeted by this attack. We observe that, regardless of the country, the majority of victims have high or very high academic qualifications and high seniority levels. This suggests that the targeted victims have the economic resources to easily match the travel pre-*payment* request from the attackers (i.e. in our case, $\approx$ 1600 USD), a request unlikely to be easily accommodated for by a junior worker or a worker with low qualifications.

It is relevant here to consider that *before* coming back at the victims, the attackers had the opportunity to inspect specific information related to the victim (including age, current position, citizenship, academic credentials, as attached to the job application). Whereas we are not in the position of evaluating replies to other targets of the attack, it is reasonable (and not unseen of [12]) that the response could be personalized over the specific characteristics of the victim. In this line, note for example that the requested advance payment is relatively modest in size for a professional in the range of those targeted by this campaign, which signals a relatively modest level of commitment required from the victim.

4) Reporting: Given the evidence collected, we attempted to contact both `SpoofedC` and `ParentC` to report the scam campaign. Unfortunately, all attempts over the phone where rejected as no internal protocol existed for such reporting. Eventually, we manged to contact the parent company `ParentC` on Twitter, who has been exceptionally responsive; once we presented the collected evidence to them, they reported this to the internal legal department. Unfortunately, having been obliged to use public channels to alert `ParentC` may have affected the effectiveness of response actions.

VI. Countermeasure effectiveness and discussion

This section provides an analysis of the effectiveness of existing countermeasures against phishing and spear phishing campaigns. A summary of the analysis is presented in Table II, where '●' indicates that the countermeasure is currently effective (i.e., research and current practices has already shown that the countermeasure can be made effective against the attack), '◑' that the countermeasure could potentially be effective (i.e., the countermeasure has the potential to counter the attack but further research is needed to find ways to make it effective), and '○' that the countermeasure does not address the criterion. The table reports a detailed rationale for each assignment.

Overall, our analysis reveals that existing countermeasures mainly focus on the characteristics of attack deployments while social engineering aspects are typically not explicitly considered. A main problem lies in the fact that technology is often unable to capture the human sphere, which plays a key role in social engineering attacks. This, however, is mainly due to an unclear formalization of social engineering vulnerabilities, attack characteristics, and attack processes that could potentially serve as an important enabler of more effective countermeasures. For example, cognitive features are mainly addressed by organizations through security-awareness training. On the other hand, the cognitive features of social engineering attacks are only *implicitly* considered in the automated detection and attack mitigation phases. For example, the ability to formally identify specific cognitive exploitation attempts in a rogue communication (e.g., an email) could enable the design and development of new phishing measurement techniques; these could be used to devise *risk metrics* for potential phishing emails, or used to evaluate the likelihood of compliance depending on the victim profile or other environmental conditions (e.g., time of day vs. attacks exploiting *scarcity*).

Our analysis suggests that the combined usage of attack characteristics and attack process provides an effective way to support the detection and mitigation of phishing attacks. However, the detection of spear-phishing attacks is more challenging due to the customization of the artifacts typically used in these attacks and the sophistication of the attack process [14]. This difficulty is created by the specific tailoring of the attack artifacts to the specific circumstances, including victim personal and technical characteristics (e.g., professional role, or local software vulnerabilities), and the ability of the attacker to *plan ahead* the steps of the attack required to achieve the final goal. For example, this allows the attacker to 'split' the cognitive profile of an attack artifact over multiple interactions as in the spear-phishing example reported above (ref. Figure 2), and using the attack structure in their favour to increase the level of commitment or trust in the victim. Similarly, *demographic* characteristics of the victims are only relevant for spear-phishing attacks, and do not play a role in 'normal' phishing due to the lack of personalization of the attack. Compare, for this purpose, the targeted professional, national, and seniority profile of victims of the reported *spear-phishing* campaign, with the general profile of the *phishing* campaign. Whereas current countermeasures ignore victim profiles, those could be useful across a number of countermeasures as highlighted in Table II.

From our analysis it emerges that a mixture of technical and human measures for the protection, detection, and response to phishing attack could be achievable. The dynamic interplay between human responses and attack phases can create a rich context for future research aimed, for example, at identifying specific 'attack signatures' that, in a risk-based fashion, can trigger an alert. These can be highly domain-dependent, whereby the appropriate countermeasure profile may change considerably depending on the type of targeted organization, and other factors such as organization culture and baseline effects. For example, natural language processing could be used to detect the nature and level of trust in a communication between entities (e.g., through sentiment analysis). This can be related backwards to previous interactions, and related to the specific cognitive characteristics of those to build a 'profile' of the attack. This, in turn, could be used to increase (or decrease) one's belief that the exchange is part of an escalating spear-phishing attack, as opposed to a 'normal' conversation.

Our analysis reveals that existing countermeasures are mostly ineffective against spear-phishing attacks as they are not able to handle the human sphere, which plays a fundamental role in those attacks. More generally, new security processes and procedures should be able to identify cognitive factors and relate them to (demographic) characteristics of potential targets. By analyzing and structuring the relationship between characteristics of social engineering attacks and defensive technologies, this paper highlights the bases on which future research can build to devise robust methods for the protection, detection, and response of advanced social engineering attacks. For example, considerations on how the attack profile may change depending on the attacker's goal (as opposed to victim selection) may provide further insights for countermeasure design and engineering.

References

[1] L. Allodi, "Economic factors of vulnerability trade and exploitation," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 1483–1499.

[2] T. Halevi, N. Memon, and O. Nov, "Spear-Phishing in the Wild: A Real-World Study of Personality, Phishing Self-Efficacy and Vulnerability to Spear-Phishing Attacks," *Social Science Research Network*, 2015.

[3] J. Wang, T. Herath, R. Chen, A. Vishwanath, and H. R. Rao, "Research Article Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email," *IEEE Transactions on Professional Communication*, vol. 55, no. 4, pp. 345–362, 2012.

[4] G. Ho, A. Cidon, L. Gavish, M. Schweighauser, V. Paxson, S. Savage, G. M. Voelker, and D. Wagner, "Detecting and characterizing lateral phishing at scale," in *28th USENIX Security Symposium*. USENIX Association, 2019, pp. 1273–1290.

[5] R. B. Cialdini, *Influence: Science and practice*. Pearson, 2009.

[6] R. Wash and M. M. Cooper, "Who provides phishing training?: Facts, stories, and people like me," in *Proceedings of Conference on Human Factors in Computing Systems*. ACM, 2018, p. 492.

[7] NDI. [Online]. Available: https://ind.nl/en/contact/Pages/Telephone-scams.aspx

[8] S. L. Blond, A. Uritesc, C. Gilbert, Z. L. Chua, P. Saxena, and E. Kirda, "A look at targeted attacks through the lense of an NGO," in *Proceedings of USENIX Security Symposium*. USENIX Association, 2014, pp. 543–558.

TABLE II
OVERVIEW OF COUNTERMEASURE EFFECTIVENESS AGAINST PHISHING AND SPEAR-PHISHING ATTACKS

Countermeasure		Phishing					Spear-Phishing				
		Scope	Soph.	Art.	Cogn.	Dem.	Scope	Soph.	Art.	Cogn.	Dem.
Protection	Training	●	●	●	◑	○	◑	◑	◑	○	◑
	Access Control	◑	○	○	○	○	○	○	○	○	○
	Inf. Prot. Procedures	◑	○	○	○	○	◑	○	○	○	◑
	Protective Tech.	◑	○	●	○	○	◑	○	○	○	○
Detection	Net. monitoring & filtering	◑	●	●	◑	○	◑	◑	○	○	○
Response	Incident reporting	○	◑	◑	○	○	○	◑	◑	○	◑
	Mitigation	◑	●	●	○	○	◑	○	◑	○	○

		Compared against	
Counterm.	Short desc.	Phishing	Spear-phishing
Training	Training to recognize and act upon social engineering attacks.	Generally effective to distinguish attacks from legitimate emails, or limit amount and nature of information available to the attacker during reconnaissance. Victim demographics usually unimportant due to untargeted nature of the attack [6].	Less effective than for phishing, as multi-stage process dilutes attack features and allows attacker to resemble ‘normal’ interactions. Demographic aspects can be relevant to personalize training (e.g., in relation to a certain risk profile for the victim).
Access Control	Set of controls to minimize the internal resources accessible by potential victims.	Generally ineffective. Can prevent part of general reconnaissance by limiting resources the attacker can probe or access to from remote [13].	Ineffective as the attacker can tailor specific attacks to the access level of the victim, and escalating from there [3], [13].
Information protocol procedures	Procedures in place to prevent the disclosure of information to attackers from employees.	Similar role as Access Control.	Attack reconnaissance can be limited by preventing detailed probing of victim systems from remote (e.g., for malware dropping), and other information useful for artifact engineering. These procedures may depend on professional roles as mobility, remote access requirements, and personal devices may play a role [8].
Protective technology	Filtering and blacklisting used to prevent access to potentially malicious external resources.	Potentially effective to detect malicious connection attempts (e.g., from low-reputation IPs during probing), and known malicious links embedded in email bodies [11], [14].	Generally less effective as in this class the components of the attack are unlikely to have been spotted before (ending up in a blacklist) [4], [8].
Network monitoring & filtering	Set of techniques and tools used to detect social engineering artifacts in transit on the network, including phishing emails, (sanitized) links, and malware.	Very effective across all dimensions, detecting malicious email bodies, scanning attempts, and known malware [11], [14]. Cognitive aspects are generally covered only indirectly in approaches employing machine learning for spam detection [9].	Much less effective due to sophistication of the engineered artifacts and the multi-stage attack process allowing the attacker to distribute the cognitive attacks across multiple interactions [4]. Spoofing and impersonation could still be a vector of detection.
Incident reporting	Incident alerting through advisories directed towards potential victims *before* the actual attack artifact(s) arrive.	Alerting does not help during reconnaissance as it can only arrive *after* the fact. It can however help detecting incoming attacks and artifacts arriving after probing.	Similar to phishing, but alerting can here be personalized over specific user characteristics (e.g., their role in the organization, or their security standing).
Mitigation	Active operations to takedown phishing domains or other resources operated by the social engineer.	Effective to limit further reconnaissance (e.g., obtaining other email addresses of possible victims), and identify and neutralize (remote) phishing resources [13].	Harder to achieve as attacks tend to disappear quickly after the one or few targeted reacted, limiting effectiveness of takedown actions and artifact identification [8], [13].

[9] R. T. Wright, M. L. Jensen, J. B. Thatcher, M. Dinger, and K. Marett, “Research note – influence techniques in phishing attacks: An examination of vulnerability and resistance,” *Information Systems Research*, vol. 25, no. 2, pp. 385–400, 2014.

[10] A. van der Heijden and L. Allodi, “Cognitive triaging of phishing attacks,” in *28th USENIX Security Symposium*. USENIX Association, 2019, pp. 1309–1326.

[11] J. Hong, “The state of phishing attacks,” *Commun. ACM*, vol. 55, no. 1, pp. 74–81, 2012.

[12] M. Paquet-Clouston, B. Haslhofer, and B. Dupont, “Ransomware payments in the bitcoin ecosystem,” *arXiv preprint arXiv:1804.04080*, 2018.

[13] F. L. Greitzer, J. R. Strozer, S. Cohen, A. P. Moore, D. Mundie, and J. Cowley, “Analysis of unintentional insider threats deriving from social engineering exploits,” in *Proceedings of Security and Privacy Workshops*. IEEE, 2014, pp. 236–250.

[14] A. Abbasi, F. Zahedi, and Y. Chen, “Impact of anti-phishing tool performance on attack success rates,” in *Proceedings of International Conference on Intelligence and Security Informatics*. IEEE, 2012, pp. 12–17.