Sonar Phishing: Pinpointing Highly Vulnerable Victims for Social Engineering Attacks

Robert Larson, Matthew Edwards, Alistair Baron, Awais Rashid

Security Lancaster
School of Computing and Communications
Lancaster University
United Kingdom, LA1 4YW

***Abstract*—Personalised attacks such as spear phishing are highly successful in compromising the security of an organisation. In this paper we propose Sonar Phishing; a technique whereby an attacker may analyse social media accounts of individuals within a company to identify victims. The victims this method selects are those whose language and actions demonstrate psychological traits indicating they are highly vulnerable to social engineering attack. We then describe how, armed with this information, attackers may generate personalised social engineering attacks, tailored to each of their targets.**

I. Introduction

Targeted social engineering attacks use personal information about an individual to create compelling behavioural hooks which draw the target to interact with a malicious payload or give out valuable information, more successfully than unsophisticated generic attacks [1]. Such attacks can hijack trust by pretending to be friends or trusted authorities [2], or can leverage a user's personal interests to entice them into responding [3], [4]. Currently, technically adept social engineers are able to craft these personalised attacks, by mining the information that the targets themselves make readily available on social networking sites (SNSs) achieving more effective phishing attacks [1].

Extracting information from SNSs and crafting personalised attacks has historically been a manual and skilled process. Whilst current tools go some of the way to allowing social engineers to perform a large-scale personalised attack across all identified employees of an organisation, carrying out an attack so broadly may undermine its effectiveness and ability to remain undetected. Automated generation of sophisticated social engineering attacks therefore still faces a number of challenges:

- Manual selection of targets: current tools allow an attacker to gather identities of individuals within an organisation. These must then be manually sifted using the judgement of the attacker, assisted by existing tools such as Maltego[1], to identify a smaller set of suitable individuals based on their appropriateness as a target [4]; such as availability of social networking content to facilitate a personalised attack.
- Identification of vulnerable victims: in attacking only a small target set, to avoid detection, it is advantageous to select those individuals most likely to be vulnerable to social engineering attacks. To identify such vulnerable individuals from their actions on SNSs requires skilled manual analysis by an attacker.
- Crafting of personalised attacks: whilst current solutions such as the Simple Phishing Toolkit[2] may harvest personal information from SNSs to generate a customised template email attacks, automate creation of profiles for impersonation purposes [5], or provide scripted interactions via a chat bot [2]; the process of creating context-aware attacks personalised to the specific vulnerabilities of an individual once again requires manual intervention by a skilled attacker.

To deal with these challenges this paper proposes the following novel solutions:

- Automated identification of highly vulnerable individuals: processing of open source content using Natural Language Processing (NLP), to identify personality traits indicating susceptibility of an individual to attack by social engineering (discussed in Section II).
- Personalised template attacks: evaluation of an individual's personality traits against a psychological attack framework, allowing a target to be attacked with the ploy to which they are most vulnerable, contextualised with personal information extracted from their SNSs (discussed in Section III).

II. Background

A. OSINT and Social Engineering

Open source intelligence (OSINT) has been shown to boost effectiveness of phishing attacks greatly. By contextualising an attack using personal information gathered from SNSs, researchers [1] found the success rate of phishing attacks rose from 16 to 72 percent.

Previous examinations of OSINT in social engineering have considered the automatic or semi-automatic extraction

[1]https://www.paterva.com/web6/products/maltego.php

[2]https://github.com/sptoolkit/sptoolkit

of artefacts from SNSs to provide the personalised context necessary to boost the success of an attack. These include:

- Enticing content: Balls et al. [4] and Brown et al. [3] propose methods to semi-automate identification of a target's interests from which to craft compelling personalised spear phishing attacks.
- Social context: Jagatic et al. [1] automate the inclusion of reassuring social context in phishing attacks, by harvesting a target's acquaintance data to spoof the origin of an email.
- Impersonation: By masquerading as the acquaintance of a target, attackers are able to leverage existing trust-relationships. Several researchers [2], [5] have identified methods of automated profile cloning that facilitate indirect attack vectors, such as friend requests and wall posts, supported by impersonation of target's relation.

B. Social Engineering Victim Psychology

Literature on the personality traits associated with social engineering victimhood is in early stages. Two key studies provide guidance on factors that may be used to predict vulnerability:

During a six month study into susceptibility to the persuasive techniques used by social engineers, Workman [6] evaluated self-reported measures of personality against objective test data generated by simulated phishing activity. Trust and obedience to authority, along with normative, continuance and affective commitment levels, were found to be significantly associated with vulnerability to attack.

Using an online questionnaire, Modic & Lea [7] examined 67 victims of scam scenarios. This study included assessment of respondents' personality traits (surgency, agreeableness, conscientiousness, emotional stability, and intellect) as represented by the Five-Factor Model [8], as well as measures of self-control and impulsivity. Here, low premeditation, low extroversion, high agreeableness and greater levels of education were found to be significant factors in predicting susceptibility to scams.

C. Estimating Psychological Traits from Social Media

Quercia et al. [9] provide a descriptive discussion of the Big Five personality traits of popular and influential users on Twitter, and also demonstrate that personality traits can be predicted with good accuracy by a regression model using only three summary measures available via Twitter — the number of users a profile is following, is followed by and has listed.

Celli [10] go further, providing an unsupervised approach to estimating Big Five personality measures, using a set of 22 correlations between linguistic features exposed in the FriendFeed social networks. They found that the personality models generated by this approach are consistent over multiple posts by the same user.

Park et al. [11] provide a strong recent investigation of the validity of estimating user profiles from social media language. Working with a large dataset of Facebook profiles, matched to self-reported personality traits using traditional questionnaire assessment, they found positive results indicating that this method is suitable for large-scale and automated analysis of personalities.

D. Tailoring social engineering attacks

Uebelacker and Quiel [12] have mapped the personality traits of victims against the principles of influence used by social engineerings, creating the Social Engineering Personality Framework (SEPF); allowing identification of the social engineering ploys an individual is most vulnerable to, based on their personality traits.

III. Automatic Identification of Highly Vulnerable Targets

By automatically revealing the most vulnerable staff members within a target organisation, for which a personalised attack may be generated, sonar phishing allows an attacker to maximise the impact of the social engineering attacks.

A. Attack model

Our proposed methodology automates the attack process, reporting an assessment of the highly vulnerable individuals that form the social engineering attack surface of a target organisation. This attack model is presented as Figure 1.

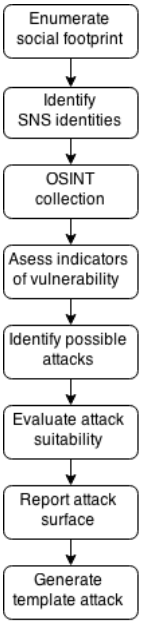

Fig. 1. Sonar phishing attack model

The initial stage of the attack identifies the online footprint of an organisation, its web-content, utilised SNSs, pages, groups, geo-location and networks. From here, the attacker may identify individuals associated with the target for OSINT collection, either through direct identification (e.g company website staff directory) or via posted SNS content and relationships.

Next, OSINT is collected from SNSs for identified individuals (e.g. relationships, likes, preferences, memberships,

and posted text). This information can be used to determine vulnerability to attack, and later to generate personalised attacks.

Collected text is then processed, using NLP techniques, for the frequency of linguistic markers that are indicators of personality traits, as identified in [10]. This allows a Five-Factor Model to be applied to each individual, a personality model that may then be compared against known indicators of vulnerability to social engineering attack, such as the low extroversion and high agreeableness identified by Modic & Lea [7]. The definitions of low and high can be considered flexible based on the size of the original target set, but could range from strictly whether the measurements fall above or below 50% to one or two standard deviations above and below the population mean.

Following vulnerability assessment, each individual's harvested OSINT is assessed against the requirements of a range of potential attacks, to evaluate which attacks are facilitated by the collected data. Available attacks are then ranked for effectiveness, by comparing the personality type of the target against the attack on the Social Engineering Personality Framework (SEPF).

The attack surface of a target organisation is then reported as a ranking of potential victims by indicators of susceptibility to attack, and the availability of the possible attacks to which they are most susceptible.

From here the attacker may select the most vulnerable individuals within a target organisation and generate a template attack, personalised using their social content, which attempts the social ploy to which they are most susceptible.

B. Example attack: targeted spear phishing of RSA

In 2011 RSA were the victim of a targeted spear phishing attack[3], that successfully delivered a malicious payload, via two emails sent to only four individuals[4] within the company. RSA speculate that this attack began with the collection from social media sites of publicly available information about specific employees, in a bid to target the attacks.

Applying the sonar phishing attack model to this attack would proceed as follows:

1) Enumerate social footprint: The online presence of RSA would be investigated using OSINT tools (e.g. Maltego) to identify the footprint of the company.
2) Identify SNS identities: SNS groups, pages etc identified in the previous stage are mined to identify individuals associated with RSA. This for example could include individuals identifying their workplace and job title on LinkedIn.
3) OSINT collection: Collection of available data from identified individuals (e.g. posted text content, group memberships, preferences). RSA identified that the attack came from a spoofed address of a supplier (Beyond), which was likely identified from OSINT, in the form of a recruitment campaign on the Beyond website (a recruitment company used by RSA).
4) Assess indicators of vulnerability: Content collected during stage three is processed for indicators of psychological traits associated with vulnerability. Here written language, such as tweets, posts, or comments would be processed using NLP techniques (as discussed in Section II) to identify linguistic indicators of vulnerability.
5) Identify possible attacks: Content collected during stage four is analysed to determine which attacks it may facilitate. Through collection of open source intelligence, 'Beyond.com' was identified as a supplier to the HR department (recruitment), providing information useful for spoofing a supplier's email address, and informing content of possible attacks.
6) Evaluate attack suitability: Available attacks ranked against susceptibility of vulnerable targets. Here members of identified HR staff may be attacked with an email from the recruitment supplier.
7) Report attack surface: Possible attacks are reported back to the attacker, ranked by vulnerability.
8) Generate template attack: Attacker generates a template attack, selecting a small number of highly vulnerable targets for attack, with an appropriate ploy. An email attack is sent to a small group of RSA HR staff, spoofing the address of their recruitment supplier. Content references a recruitment plan, and instructs the HR staff to open the malicious payload attached to the email.

[3]https://blogs.rsa.com/anatomy-of-an-attack/

[4]http://www.wired.com/2011/08/how-rsa-got-hacked/

IV. Conclusions

In this paper, we introduce sonar phishing: a method of greatly increasing the effectiveness of phishing by automatic selection of highly vulnerable victims for social engineering. We propose that by combining existing techniques of open-source intelligence (OSINT) collection and natural language processing, attackers may mine social media content to find victims. They can do this by observing indicators of psychological traits that current literature suggests are associated with susceptibility to social engineering attacks. Armed with this data we propose a method to deliver automatically-generated template attacks containing ploys personalised to the specific psychological weaknesses of an individual within a target organisation.

Our theoretical attack process is based on the combination of literature from numerous fields, and therefore further work is needed to produce a prototype system for evaluation of the proposed sonar phishing process. Here we have discussed current methods of OSINT collection and highlighted literature that identifies personality traits linked to susceptibility to social engineering, however a framework mapping these to OSINT data sources is needed. So too is a taxonomy of social engineering ploys, mapping the OSINT requirements of possible social engineering attacks to allow for automated generation of attacks.

Our approach highlights that the vulnerability of an organisation to targeted social engineering attack results from a

combination of the psychological susceptibility of individuals within an organisation, and the availability of open source information. To mitigate this threat an organisation could carry out the initial stages of a sonar phishing attack against itself, to identify problematic SNS content, highlight vulnerable staff who may require security awareness training, or flag individuals in breach of company social media usage policy.

Personalised targeted attacks such as those generated by our proposed approach have proven highly effective and difficult to detect with current approaches. The method proposed here to automate attack crafting identifies the psychological foundations of the social engineering ploys used in attacks. In order to mitigate these threats, further work is necessary to identify these sophisticated attacks by the linguistic markers of the psychological ploys they contain.

REFERENCES

[1] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, Oct. 2007. [Online]. Available: http://doi.acm.org/10.1145/1290958.1290968

[2] M. Huber, S. Kowalski, M. Nohlberg, and S. Tjoa, "Towards automating social engineering using social networking sites," in *Proceedings IEEE CSE'09, 12th IEEE International Conference on Computational Science and Engineering, August 29-31, 2009, Vancouver, BC, Canada*, 2009, pp. 117–124. [Online]. Available: http://dx.doi.org/10.1109/CSE.2009.205

[3] G. Brown, T. Howe, M. Ihbe, A. Prakash, and K. Borders, "Social networks and context-aware spam," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 2008, pp. 403–412.

[4] L. Ball, G. Ewan, and N. Coull, "Undermining - social engineering using open source intelligence gathering," in *KDIR 2012 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, Barcelona, Spain, 4 - 7 October,2012*, 2012, pp. 275–280.

[5] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: Automated identity theft attacks on social networks," in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 551–560. [Online]. Available: http://doi.acm.org/10.1145/1526709.1526784

[6] M. Workman, "Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 4, pp. 662–674, 2008.

[7] D. Modic and S. E. Lea, "How neurotic are scam victims, really? the big five and internet scams," in *Conference of the International Confederation for the Advancement of Behavioral Economics and Economic Psychology, Exeter, United Kingdom*, 2011.

[8] L. R. Goldberg, "Language and individual differences: The search for universals in personality lexicons," *Review of personality and social psychology*, vol. 2, no. 1, pp. 141–165, 1981.

[9] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our twitter profiles, our selves: Predicting personality with twitter," in *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*. IEEE, 2011, pp. 180–185.

[10] F. Celli, "Unsupervised personality recognition for social network sites," in *ICDS 2012, The Sixth International Conference on Digital Society*, 2012, pp. 59–62.

[11] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. Seligman, "Automatic personality assessment through social media language." *Journal of Personality and Social Psychology*, 2014.

[12] S. Uebelacker and S. Quiel, "The social engineering personality framework," in *2014 Workshop on Socio-Technical Aspects in Security and Trust, STAST 2014, Vienna, Austria, July 18, 2014*, 2014, pp. 24–30. [Online]. Available: http://dx.doi.org/10.1109/STAST.2014.12