(Poster) 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
email: r.larson,m.edwards7,a.baron,marash@lancaster.ac.uk

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]. 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 (up to 52% increase [1]).

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

- **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, to identify a smaller set of suitable individuals based on their appropriateness as a target [3], 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. To date, there is no method for an attacker to automate this process, therefore an unskilled attacker may inadvertently target those less susceptible to social engineering, increasing the risk of detection.
- **Crafting of personalised attacks**: While current solutions such as the Simple Phishing Toolkit may harvest personal information from SNSs to generate customised template email attacks, or automate the creation of profiles for impersonation purposes [4], the process of creating context-aware attacks personalised to the specific vulnerabilities of an individual requires manual intervention by a skilled attacker.

We propose Sonar Phishing, a novel attack model that demonstrates how research from the fields of natural language processing and psychology, make it feasible for an attacker to overcome the following restrictions of automating highly targeted social engineering attacks:

- **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.
- **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 SNSs.

Current defensive solutions to social engineering focus passively on security awareness training and hardening of operational procedures, with limited methods to evaluate their effectiveness. The Sonar Phishing approach, employed as a penetration testing tool, would allow an organisation to proactively audit its social engineering attack surface.

II. Components of Sonar Phishing attack model

Sonar Phishing provides valuable information to the attacker, regarding the social engineering attack surface of an organisation at each stage of the attack. It processes natural language content, extracted from SNSs to automatically identify the most vulnerable individuals in an organisation, and culminates in a spear phishing attack, tailored to the specific vulnerability of the individual. The attack has four main components:

A. Collection of Open Source Intelligence (OSINT)

OSINT gathered from SNSs, used to create personal context to a phishing attack have been shown in existing research, to increase the rate of success from 16 to 72% [1]. Methods used include: enticing content (e.g., interests) [3], and leveraging of trust-relationships (e.g., profile cloning) [2].

By creating a taxonomy of the OSINT requirements of social engineering attacks, for comparison with collected data, the proposed Sonar Phishing model allows us to identify which attacks may be performed against a target, using their own data.

B. Identifying Psychological Traits from Social Media

Identification of psychological traits from SNSs can be achieved through correlation of linguistic factors in SNS content with the personality traits of a user. Several large scale studies [5], [6] combine these two factors to form a personality model. Examples of linguistic factors that correlate with aspects of personality identified by these studies [7], [8] are: topic (e.g. subject of discussion), word usage (e.g. frequency of word use), and psychological aspects (e.g. textural expression of feelings).

Our proposed process model harnesses these methods to generate profiles of the psychological traits of individuals within an organisation from their gathered OSINT.

C. Identifying Vulnerable Targets

Two key studies [9], [10] identify the personality traits that may act as predictors of vulnerability to social engineering attacks, including: low premeditation, low extroversion, high agreeableness.

By examination of identified personality models for indicators of the traits associated with vulnerability to social engineering, our approach is able to evaluate an individual's vulnerability to attack.

D. Tailoring Social Engineering Attacks

Uebelacker and Quiel [11] have mapped the personality traits of victims against the principles of influence used by social engineers, creating the Social Engineering Personality Framework (SEPF). By leveraging the SEPF against indicators of an individual's personality traits, gathered during OSINT collection, Sonar Phishing is able to automatically identify the attack to which a target is most vulnerable.

III. Stages Of Fully-Automated Attack

Unlike current solutions, Sonar Phishing provides total automation of the attack process, allowing an unskilled attacker to carry out highly-targeted sophisticated social engineering attacks against individuals in a target organisation. The key stages are:

1) **Enumerate social footprint:** identity the online footprint of an organisation (web-content, SNSs, groups etc).
2) **Resolve SNS profiles:** identify individuals associated with the target for OSINT collection (e.g., SNS group membership, profile information).
3) **OSINT collection:** OSINT is collected from SNS profiles for identified individuals (e.g., relationships, preferences, memberships, and posted text).
4) **Generate profile:** collected text is then processed, using NLP techniques, for the frequency of linguistic markers that are indicators of personality traits, generating a five-factor model[7] for each individual.
5) **Assess indicators of vulnerability:** personality models are assessed against known indicators of vulnerability to social engineering attack.
6) **Evaluate available attacks:** harvested OSINT is assessed against an attack taxonomy to determine which attacks can be performed with the collected data.
7) **Rank attack effectiveness:** available attacks are ranked for effectiveness, by comparing the personality model of the target against the attack on the Social Engineering Personality Framework (SEPF).
8) **Report attack surface:** a ranking of potential individuals, by indicators of susceptibility to attack, and the availability of the possible attacks to which they are most susceptible, is presented.
9) **Generate attack:** Launch a template social engineering ploy, of the type to which the target is most vulnerable (ranked on SEPF), personalised to the individual with their own SNS content.

IV. Future Work

Future work will focus on implementation of a toolkit to operationalise modules of existing research and complete the Sonar Phishing approach. Key to this process will be creation of a detailed taxonomy of social engineering attacks, mapping ploys to their required data. This will provide insight into how to better detect and defend against such threats, support staff security awareness training, and the enhancement of security procedures. Employed as a penetration testing tool, Sonar Phishing allows an organisation to evaluate the effectiveness of these efforts, without an experienced social engineer. Our approach facilitates this, through automatic identification of SNS content causing an OSINT threat, and automated auditing of employee vulnerability to attack.

References

[1] T. N. Jagatic *et al.*, "Social phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, Oct. 2007.

[2] M. Huber *et al.*, "Towards automating social engineering using social networking sites," in *Proc. IEEE CSE*, 2009, pp. 117–124.

[3] L. Ball, G. Ewan, and N. Coull, "Undermining - social engineering using open source intelligence gathering," in *KDIR*, 2012, pp. 275–280.

[4] L. Bilge *et al.*, "All your contacts are belong to us: Automated identity theft attacks on social networks," in *Proc. WWW*. NY, USA: ACM, 2009, pp. 551–560.

[5] F. Mairesse *et al.*, "Words mark the nerds: computational models of personality recognition through language," in *Proc. COGSCI*, 2006, pp. 543–548.

[6] D. Quercia *et al.*, "Our twitter profiles, our selves: Predicting personality with twitter," in *Proc. PASSAT and IEEE SocialCom*, 2011, pp. 180–185.

[7] F. Celli, "Unsupervised personality recognition for social network sites," in *Conf. ICDS*, 2012, pp. 59–62.

[8] L. Qiu *et al.*, "You are what you tweet: Personality expression and perception on twitter," *Journal of Research in Personality*, vol. 46, no. 6, pp. 710 – 718, 2012.

[9] M. Workman, "Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security," *JASIST*, vol. 59, no. 4, pp. 662–674, 2008.

[10] D. Modic and S. E. Lea, "How neurotic are scam victims, really? the big five and internet scams," in *Conf. ICABEEP*, 2011.

[11] S. Uebelacker and S. Quiel, "The social engineering personality framework," in *Workshop on STAST*, 2014, pp. 24–30.