Undermining
Social Engineering using Open Source Intelligence Gathering

Leslie Ball, Gavin Ewan and Natalie Coull
School of Engineering, Computing and Applied Mathematics, University of Abertay Dundee, Bell Street, Dundee, U.K.

Keywords: Social Engineering, Cognitive Hacking, Open Source Intelligence, Phishing, Data Mining, Visualisation.

Abstract: Digital deposits are undergoing exponential growth. These may in turn be exploited to support cyber security initiatives through open source intelligence gathering. Open source intelligence itself is a double-edged sword as the data may be harnessed not only by intelligence services to counter cyber-crime and terrorist activity but also by the perpetrator of criminal activity who use them to socially engineer online activity and undermine their victims. Our preliminary case study shows how the security of any company can be surreptitiously compromised by covertly gathering the open source personal data of the company's employees and exploiting these in a cyber attack. Our method uses tools that can search, drill down and visualise open source intelligence structurally. It then exploits these data to organise creative spear phishing attacks on the unsuspecting victims who unknowingly activate the malware necessary to compromise the company's computer systems. The entire process is the covert and virtual equivalent of overtly stealing someone's password 'over the shoulder'. A more sophisticated development of this case study will provide a seamless sequence of interoperable computing processes from the initial gathering of employee names to the successful penetration of security measures.

1 INTRODUCTION

The ubiquity of online social networking sites and the blogosphere provides a reservoir of relatively untapped data in terms of extracting value from their analysis. Many opportunities are thus presented by these data. Specific to security, the data may act as open source intelligence (OSINT) for the benefit of intelligence services and military strategists to counter cyber-crime and terrorist activity. Indeed, effective counter-terrorism has been deemed unsuccessful "without the adequate exploitation of open source information" (Borchgrave et al. 2006). Other work has integrated OSINT into the wider intelligence cycle in terms of crowd sourcing and the empowerment of the public (Steele, 2007). More broadly, automatic analyses have been made on other forms of open source information such as the natural language processing of social media to model the prediction of social tension (Vybornova et al., 2011) and to detect emergent conflict through web mining (Johansson et al., 2011). These are only a few examples of literature that report on the benefits of social media analysis to those involved in counter-terrorism, military strategy and social and political stability. Contrary to the societal benefits of OSINT, it can also be harnessed by those intent on perpetrating crime. In the remainder of this paper we focus on this issue and, in particular, on the coupled and covert processes of data mining and social engineering employed by the cyber-criminal to deceive the user into disclosing security data for access to computer systems.

The next section introduces the phenomenon of social engineering and how it can be a very effective tool for breaching security and complementary to the technical hacking techniques, which are a direct attack on a computer system and do not involve the behavioural aspects of the user. Thereafter we present a preliminary case study, which illustrates the effectiveness of a covert process of intelligence gathering integrated into social engineering to compromise computer security.

2 SOCIAL ENGINEERING

Social engineering is "the art of gaining access to buildings, systems or data by exploiting human psychology, rather than by breaking in or using technical hacking techniques" (CSO Magazine, 2012). As technology becomes more sophisticated

Ball L., Ewan G. and Coull N..
Undermining - Social Engineering using Open Source Intelligence Gathering.
DOI: 10.5220/0004168802750280
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2012), pages 275-280
ISBN: 978-989-8565-29-7

and users more inter-connected as a result, virtual social interaction has inevitably followed on a huge scale. Facebook accounts for approximately 3 in 4 minutes spent on social networking sites and 1 in every 7 minutes spent online around the world (The New Age, 2011). Moreover, digital deposits now go far beyond the superficial "toast and coffee for breakfast" type of blog. Website forums, Facebook, Twitter, Tumblr, Wordpress.com are all examples of social networking media that mobilise social networks and allow the expression of opinions and the disclosure of personal data. These data are, to varying degrees, public and therefore open to exploitation.

Our inter-connected virtual society has presented opportunities to the malicious hacker, not only in terms of direct brute force attack but also in terms of psychological manipulation, and both contribute towards what is known as the *vector attack* in computer security terminology. Security of Information Technology has thus become a major concern for companies and governments. In 2010 in the UK, cyber-terrorism was prioritised as a Tier One threat to national security by the government. The term cyber security is widely adopted to define this phenomenon.

In the literature, the terms *social engineering* and *cognitive hacking* appear to be synonymous, though the latter has appeared less recently since it was coined within a body of work by Cybenko et al. (2002) and Giani and Thompson (2007). Enrici et al. (2010) offer a discourse on the cognitive profiling of a computer hacker and the psychological effects of human factors in terms of usability and of human errors in terms of failure, all within the context of IT security.

Stech (2011) confirms that there have been few publications that map the social and behavioural aspects of cyber-deception to the classical denial and deception tactics adopted in conventional warfare. Rather, the focus has been on recognising that a social engineering attack incorporates both technical and social considerations that feed on the lethargy of the user regarding security and the aggression of the malicious hacker (Abraham and Chengalur-Smith, 2010). This combination is further endorsed by Maan and Sharma (2012). A framework of feedback loops has also been considered to model the manoeuvres of the attacker against those of the organisational countermeasures, where they postulate that an organisation's technical defences are superior to their human equivalents (Gonzalez et al., 2006). The same authors argue that the key for the social engineer is to make the countermeasures transparent so that they can be incorporated into the main attack feedback loop, which measures the outcomes of each attack, in order to evaluate the next action to take.

With specific reference to social media content, the use of natural language processing has been used to measure information assurance (Raskin et al., 2010). This technique applies to monitoring suspicious activity at social networking sites, where postings may exhibit inconsistency and therefore expose the possibility of uncovering insider threats to social engineering attacks. Linked to this is research implementing an automated social engineering bot attack on social media sites such as Twitter ad Facebook (Huber et al., 2009). In a recent review, Heikkinen (2010) states how the user can be lulled into a false sense of security knowing that the company implement firewall strategies and virus detection, and emphasise the importance of user training. The focus of our paper encapsulates the spirit of Heikkinen's work as well as encompassing the notion of the partial technical and social attack of other authors' research already outlined.

The next section presents our case study to illustrate the creative ideas behind the processes of social engineering to compromise security measures on a computer system.

3 CASE STUDY

The focus of the case study is on the proposed attack of a company with whom we have previously consulted. For privacy, we refer to the company as X hereafter. The key to unlocking the security measures on X's computer system is its employees by exposing them to a vector attack. All employee data have also been made anonymous. The full process of how the employees may be deceived to disclose the necessary information to breach security is revealed.

3.1 Aim

The purpose of the case study is to demonstrate show how a malicious attacker, coupled with the appropriate use of software tools can harness and integrate open intelligence gathering into the social engineering process to bring about a successful vector attack.

3.2 The Procedure

We adopt a sequence of events to illustrate how a

socially engineered vector attack can be organised around a target. The plan of attack is as follows and encompasses the key stages of searching for employee profiles, drilling down for employee interests and targeting spear phishing attacks:

1. Prepare the software platform to perform the strategic searches.
2. Implement a covert search on company employees and extract detail of interest.
3. Construct a spear phishing attack targeted at vulnerable employees.
4. Propose countermeasures to the social engineering process.

Each of the stages is considered in the remainder of this subsection. The first three stages illustrate the attack, while the last stage considers how a company might counter such an attack in the real world.

3.2.1 Acquisition of Tools

Software tools were used to effectively search company X's website and affiliated online content. The tools were able to extract employee names, identify some personal interests of these employees and to craft a spear phishing attack based on these data. We used Maltego to perform the first two data gathering exercises and the Simple Phishing Toolkit to construct the email attack. Maltego is an open source intelligence and forensics application program, which is capable of mining internet websites, Twitter feeds and other social media content. The Simple Phishing Toolkit is a relatively new addition to the social engineer's arsenal and allows the construction of a spear phishing campaign, which is essentially the confidence trick of the operation. Most importantly, the tool provides a website 'scraper' facility. This facility can effectively extract details from a website that is deemed of interest to a targeted individual and design an email template that appears to have been sent from this website. This is the crucial stage of deception.

3.2.2 Covert Intelligence Gathering

This stage proceeds with the processes of intelligence gathering. At this prototype stage the entire sequence of events is not automated and requires intermittent manual intervention. Company X's website contains a list of employee names and each member of staff displays various degrees of personal and work-related information. We selected as targets only those employees who displayed detailed information about themselves (*i.e.* both personal and work information). The vetted list was then input to Maltego, using manual intervention. Figure 1, shows the full anonymised list of staff email addresses.

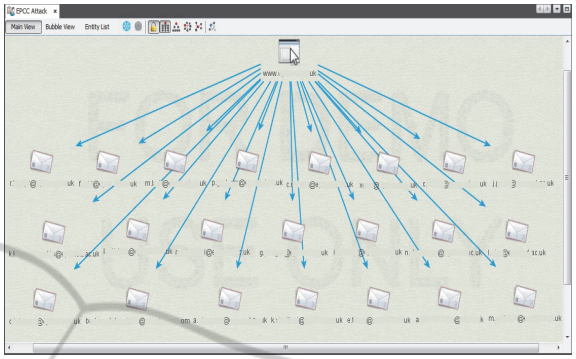

Figure 1: The initial search on company X's website extracts employee emails (anonymised).

Having assimilated a target list of employees the next stage was to run what are termed *transforms*. These perform the drilling down process and there are many options from which to choose. This case study made the assumption that other emails associated with the employees in Figure 1 would be a good transform to perform. While the results for this transform are potentially large, Figure 2 illustrates the returned information for one of the employees only.

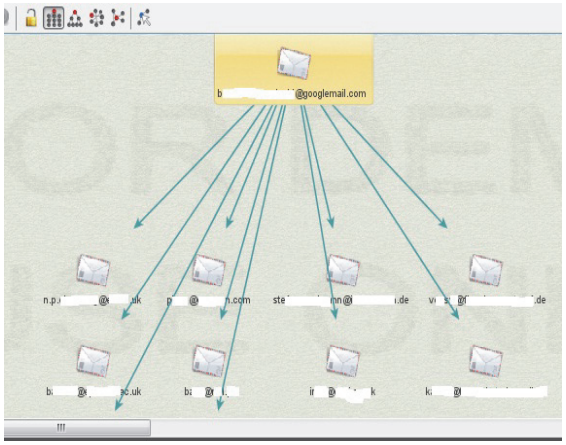

Figure 2: The anonymised transform results for emails associated with a single employee.

While the bulk of these returns did not show anything much of interest, other transforms based on common interests did. With this type of transform we found that many employees linked to the same interest such as, for example, hill walking. This information is invaluable to the social engineer planning a vector attack. Maltego offers a

visualisation module to enhance the display of complex data. Using what is termed an 'edge-weighted view' the visualisation in Figure 3 shows which of the employees link to which common interest. The more links to an interest the bigger the bubble representation. Hill walking is seen as the most common interest in this group of employees, followed by Badminton and Travel. These items are those that the social engineer will use to plan their vector attacks.

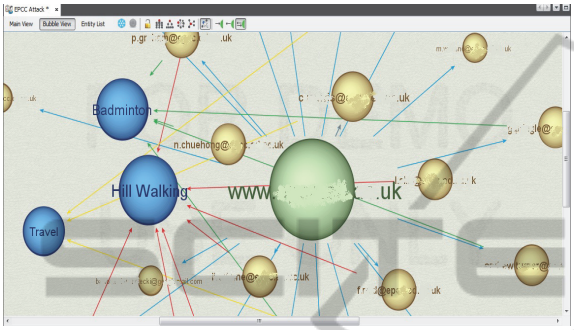

Figure 3: The bubble representation that links employees (anonymised) to common interests.

Armed with these data, the social engineer can now proceed to the next stage of designing the spear phishing attack, which could specifically target, say, all those employees interested in hill walking.

3.2.3 The Vector Attack

The objective of the spear phishing attacks is to use the covertly gathered information from previous processes described in order to increase the likelihood of successfully penetrating the security system offsite.

Phishing is a confidence trick to steal personal information, usually by email and, whilst it is not a new phenomenon, spear phishing is a relatively recent tool addition for the malicious attacker. Though email phishing attacks still succeed, we are generally much more aware in terms of recognising these types of attack as they adopt typical designs such as a generic greeting, a sense of urgency or a direct request for personal information. Furthermore, automated spam filters use these same criteria to intercept suspicious looking emails. However, the same measures fall short when encountering a spear phishing attack. This is because, unlike generic phishing attacks, they are not issued widely and randomly but rather they target individuals and so adopt a more socially aware design into their emails. The goal is, however, the same in that they will ask the receiver to click a link, which may indeed appear to be urgent, though related to their interests directly. It is the contextualising of the request within a personalised email that increases the probability of success. By targeting the individual in this way the attack seeks to abuse the relationship of trust and this falls categorically into the repertoire of the social engineer (Williams, 2011).

Using this approach, the generic emails of phishing attacks have essentially been replaced with emails from seemingly trusted sources and by displaying the recipient's name as part of the personalisation. The spear phisher thrives on familiarity by knowing your name, your address and a little about your personal interests (Norton, no date). So, while users are now more educated on phishing attacks, how likely are they to not click on a link that has come from an apparently trusted source such as a friend or a website used by the user for leisure activity?

The Simple Phishing Toolkit allows the social engineer to construct a list of target individuals including their name, email address and groups them into categories related to their interests as in Figure 4.

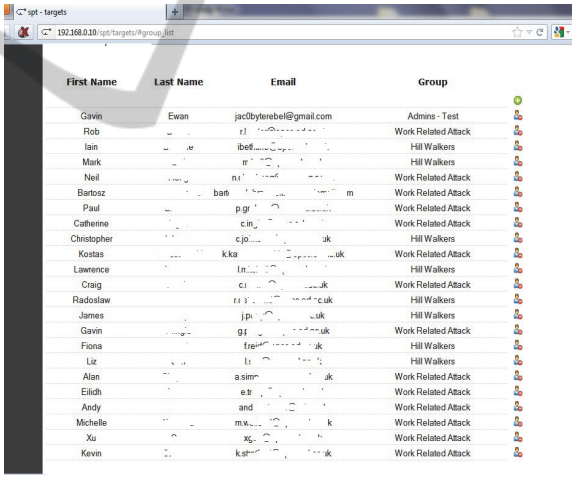

Figure 4: The Simple Phishing Toolkit generates lists of individuals to target (anonymised).

The software can then generate a personalised email for each of the groups by using the text associated with an identified website template. The flavour of the email is personal and friendly and invites the targeted individual to visit their website, a bogus website, for further information (Figure 5). Once the individual has clicked the linked to the website malware is activated that can steal security by keystroke monitoring or from user activity on Company X's server. In this respect there is never a request to the individual to disclose their security

information directly as the malware achieves this in lieu of the request. The victim has thus been undermined.

Figure 5: The spear phishing attack takes the form of a personalised email inviting the recipient to visit a website for further information (anonymised).

3.2.4 Proposal for Countermeasures

So what could any company or other venture do to secure their computing infrastructure from these types of creative vector attacks?

Firstly, it is clear from this initial case study that those staff members leaking professional and personal information are more vulnerable to a social engineering attack. A company should therefore provide the necessary education to its staff on the security threats (Abraham and Chengalur-Smith, 2010; Heikkinen, 2010) that can be engineered from public data and to always question requests to click on links that were unexpected or unsolicited. While this study only simulated an attack on those interested in hill walking, any other interest highlighted in Figure 3 could have been used to similar effect. Even less suspicious would be those email requests that are work related and seem to follow the natural course of everyday working life rather than specific to personal interests.

Taken further, a company could design and implement policies that prevent their staff from posting personal details. Without such a policy in place, the 'humanising' effect ensues, which plays straight into the hands of the social engineer who is studying the psychological behaviour of its targets in order to mimic them in the attack.

Lastly the company could take a more aggressive approach by actively spear phishing their employees explicitly in a harmless attack in order to test their awareness. It in effect becomes the company drill of a cyber-attack as a preventative measure rather than a fire drill exercise, for example.

4 CONCLUSIONS

This paper has focused on the creative process of cyber-attacks using the surreptitious techniques of social engineering. Ultimately, however, the social engineering attack has been identified as the top information security threat in 2012 (Trend Micro, 2012). A case study was designed to simulate such an attack on company X's computing infrastructure in order to highlight the vulnerability of disclosing too much data on publicly available websites.

Our spear phishing email demonstrates how using appropriate search and mining tools and manual interventions, a company's computer security could be compromised by exploiting personal details posted by the company's staff members.

By extracting an initial list of staff names and their associated interests it was possible, using open source intelligence and phishing software, to craft a personalised email that engendered trust in the user but was in fact a confidence trick to get the user to click on a link to a bogus website. By clicking on a malicious link, malware can be easily downloaded to the victim's machine, in spite of existing security measures such as firewalls and anti-virus, which could steal user credentials and other valuable information.

The work illustrates that spear phishing as opposed to normal phishing, is likely to be much more effective as they target the individual in a more socially aware design than the latter, which issues a random and blanket email attack.

The case study illustrates the various stages required to search, extract and visualise intelligence data, which are invaluable to designing the spear phishing attack. The future requirement from this work is to devise an intelligent bot that can perform the entire sequence of events seamlessly without the manual intervention of the attacker. The achievement of this with a bot would require the integration of pattern recognition algorithms for text analytics as well as decision-making capabilities on who to target and how to automate the email attack effectively.

REFERENCES

Abraham, S. and Chengalur-Smith, I., 2010. An Overview

of Social Engineering Malware: Trends, Tactics, and Implications. *Technology in Society*, 32(3): 183-196.

Borchgrave de, A., Sanderson, T. and MacGaffin J., 2006. Open Source Information: The Missing Dimension of Intelligence. *Report of the CSIS Transnational Threats Project*.

CSO Magazine, 2012. *The Ultimate Guide to Social Engineering.* [Online] Accessed 12/06/2012 at http://assets.csoonline.com/documents/cache/pdfs/Social-Engineering-Ultimate-Guide.pdf

Cybenko, G., Giani, A. and Thompson, P., 2002. Cognitive Hacking: A Battle for the Mind. *IEEE Computer*, 35(8), 50-56.

Enrici, I., Ancilli, M. and Lioy, A., 2010. A Psychological Aproach to Information Technology Security. In *3rd Conference on Human System Interaction,* 459-466.

Giani and P. Thompson. Detecting Deception in the Context of Web 2.0. In *Web 2.0 Security & Privacy, 2007.*

Gonzalez, J., Sarriegi, J. and Gurrutxaga, A., 2006. A Framework for Conceptualizing Social Engineering Attacks. *CRITIS 2006,* LNCS 4347, 79-90.

Heikkinen, S., 2010. Social Engineering in the World of Emerging Communication Technologies. In *Proceedings of Wireless World Research Forum meeting #17, Nov 2006.*

Huber, M., Kowalski, S., Nohlberg, M. and Tjoa, S., 2009. Towards Automating Social Engineering Using Social Networking Sites. In *International Conference on Computational Science and Engineering*, 3:117-124.

Johansson, F., Brynielsson, J., Hörling, P., Malm, M., Mårtenson, C., Truvé, S. and Rosell, M., 2011. Detecting Emergent Conflicts Through Web Mining and Visualization. In *European Intelligence and Security Informatics Conference 2011*, 346-353.

Maan, P. and Sharma, M., 2012. Social Engineering: A Partial Technical Attack. *International Journal of Computer Science Issues,* 9(2), 1694-0814.

The New Age, 2011. *Social Networking is the most Popular Online Activity*. [Online] Accessed 12/6/2012 at http://www.thenewage.co.za/38836-1021-53-Social_networking_is_the_most_popular_online_activity

Norton., [no date]. *Spear Phishing: Scam, not Sport*. [Online] Accessed 12/06/2012 at http://uk.norton.com/spear-phishing-scam-not-sport/article

Raskin, V., Taylor, J. and Hempelmann, C., 2010. Ontological Semantic Technology for Detecting Insider Threat and Social Engineering. *NSPW'10*, 21-23 Sept. 2010, Concord, MA, 115-127.

Stech, F., Heckman, K., Hilliard, P. and Ball, R., 2011. Scientometrics of Deception, Counter-deception, and Deception Detection in Cyber-space. *PsychoNology Journal*, 9(2), 79-122.

Steele, R., 2007. Open Source Intelligence. In Johnson, L. (ed.) *Strategic Intelligence: The Intelligence Cycle*, Praeger. Westport CT, 96-122.

Trend Micro, 2012. *Social Engineering Remains Top Security Threat in 2012*. [Online] Accessed 12/06/2012 at http://www.newswit.com/.it/2012-04-05/bf9543225f9137e29c7a64af58a75c2b/

Vybornova, O., Smirnov, I., Sochenkov, I., Kiselyov, A. and Tikhomirov, I., 2011. Social Tension Detection and Intention Recognition Using Natural Language Semantic Analysis. *European Intelligence and Security informatics Conference 2011*, 277-281.

Williams, C., 2011. *Google Cyber Attacks: What is Spear Phishing?* [Online] Accessed 12/06/2012 at http://www.telegraph.co.uk/technology/news/8552297/Google-cyber-attacks-what-is-spear-phishing.html