Panning for Gold: Automatically Analysing Social Engineering Attack Surfaces

Matthew Edwards*, Robert Larson, Benjamin Green, Awais Rashid, Alistair Baron*

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

Abstract

The process of social engineering targets people rather than IT infrastructure. Attackers use deceptive ploys to create compelling behavioural and cosmetic hooks, which in turn lead a target to disclose sensitive information or to interact with a malicious payload. The creation of such hooks requires background information on targets. Individuals are increasingly releasing information about themselves online, particularly on social networks. Though existing research has demonstrated the social engineering risks posed by such open source intelligence, this has been accomplished either through resource-intensive manual analysis or via interactive information harvesting techniques. As manual analysis of large-scale online information is impractical, and interactive methods risk alerting the target, alternatives are desirable.

In this paper, we demonstrate that key information pertinent to social engineering attacks on organisations can be passively harvested on a large-scale in an automated fashion. We address two key problems. We demonstrate that it is possible to automatically identify employees of an organisation using only information which is visible to a remote attacker as a member of the public. Secondly, we show that, once identified, employee profiles can be linked across multiple online social networks to harvest additional information pertinent to successful social engineering attacks. We further demonstrate our approach through analysis of the *social engineering attack surface* of real critical infrastructure organisations. Based on our analysis we propose a set of countermeasures including an automated social engineering vulnerability scanner that organisations can use to analyse their exposure to potential social engineering attacks arising from open source intelligence.

Keywords: Social Engineering, Vulnerability Analysis, Open Source Intelligence, Social Networks, Competitive Intelligence

1. Introduction

Social engineering attacks pose a major risk to the security of organisations. Some of the most high profile cyber attacks on large organisations, e.g., RSA, JP Morgan, AT&T, the Ukranian power grid, etc., leveraged social engineering as an entry point into the organisation's systems. Attackers use a number of tactics, ranging from simple impersonation to complex multi-layered deceptions worthy of a Hollywood caper, that lead a target to disclose sensitive information or to interact with a malicious payload. At their most basic, these attacks may be represented by a generic phishing email from an unfamiliar sender that targets hundreds of staff within an organisation with the same message. More sophisticated attacks may greatly increase their chance of success by targeting a much smaller pool of recipients with a personalised ploy [1].

Current research suggests that the effectiveness of such attacks can be greatly increased through the use of open source intelligence (OSINT) to boost the effectiveness of the deceptive ploys delivered in an attack [1]. Such open source information is now widely available – with individuals increasingly releasing information about themselves online, particularly on social networks. Even more worryingly, practices such as organisational engagement with social media and the publication of employee rosters on organisational websites are enabling attackers to easily identify an organisation's employees from amongst millions of social media users. This lets attackers know exactly who to target for data harvesting in preparation for an attack on the organisation. Methods by which such OSINT data may be used to increase effectiveness in this manner include (but are not limited to): selection of vulnerable personalities, inclusion of ploys personally attractive to the target, and impersonation of a person in authority [2].

Existing research has demonstrated the social engineering risks posed by such OSINT data [3]. However, this normally relies on labour intensive manual analysis [4], which is impractical and poses a high cost to a potential attacker. Alternatively, such techniques utilise automated conversational agents [2], which do not scale and are not very effective due to the challenges of imitating human conversational behaviour. Other techniques rely

*Corresponding authors

Email addresses: m.edwards7@lancaster.ac.uk (Matthew Edwards), r.larson@lancaster.ac.uk (Robert Larson), b.green2@lancaster.ac.uk (Benjamin Green), a.rashid@lancaster.ac.uk (Awais Rashid), a.baron@lancaster.ac.uk (Alistair Baron)

on "active" engagement with potential targets to elicit information — through zombie profiles or misleading friend requests [5] — and hence risk detection prior to an attack being launched. In this paper, we demonstrate that both of these challenges – automation and passive information gathering – can be overcome, posing major social engineering risks to organisations.

We show that it is possible to automatically identify the employees of an organisation amongst individuals within its online footprint. Furthermore, we demonstrate that it is possible to automatically resolve employee identities across multiple online social networks, with a high accuracy, for large-scale harvesting of information pertinent to launching social engineering attacks. We also show that such harvesting can be undertaken "passively" without resorting to invasive measures, enabling vulnerability assessments which do not rely on exercising deception during social engineering penetration tests. Through automated identification of OSINT that may be used to conduct or enhance a social engineering attack against an organisation, we aim to highlight potential risks to the target, allowing appropriate mitigation techniques to be selected.

The key contributions of our work are as follows:

- In-depth interviews were conducted with expert social engineering penetration testers to better understand the variety of social engineering attacks used, and how OSINT data facilitates the attacks. A summary of the valuable insights from these interviews is presented in Section 3.
- We present an automated approach for identifying the employees of an organisation from amongst the many connected profiles in online social networks. So far as we are aware, no previous work exists on the topic of automatically identifying — from only public data — which of an organisation's social media followers are actually its employees. The nearest approximation we are aware of is Scheelen et al. [5], who investigated a single company by connecting with followers on LinkedIn, where the social media structure is based around employment.
- We present an approach for automated resolution of identities across social media – demonstrating that large-scale harvesting of such information is feasible for attackers. While employees may be careful about their presentation in online profiles linked to their work identity, we identify features that can be used to link profiles on different online social networks. We present an ensemble classifier, which makes its decision about whether two profiles can be matched based on the reported matches of sub-classifiers working on specific profile features. While more advanced methods exist which could produce more accurate comparison results for each feature, we employ unsupervised methods which release us from the requirement of obtaining training data for the subclassifiers and which are relatively computationally inexpensive.
- We provide an analysis of the online footprints of 13 critical infrastructure companies, demonstrating the extent of their vulnerability to social engineering attacks based on employee information in online social media. We discover that material sufficient to launch sophisticated email and phone attacks targeted at employees is automatically reachable for all but one of the examined organisations.
- We propose a number of mitigation strategies and make our approach — an automatic social engineering vulnerability scanner — available for organisations to counter such risks[1].

The rest of this paper is structured as follows. In Section 2 we discuss related work connecting OSINT and social engineering. In Section 3 we summarise the findings from in-depth interviews with social engineering professionals. In Section 4.1 we demonstrate how automated methods can be deployed to identify a company's employees from amongst its followers on Twitter, while in Section 4.2 we detail and evaluate our probabilistic identity resolution system on profiles from across four major online social networks (OSNs). In Section 5 we go on to present the results of automated analysis on the digital footprints of critical infrastructure organisations. Section 6 presents the final product of the research as a vulnerability scanner and mitigation tool, evaluating its performance with five companies. In Section 7 we discuss our results and reflect on the implications for social engineering penetration testing and organisational practices for online security. We draw conclusions and offer suggestions for future work in Section 8.

2. Related Work

2.1. OSINT and Social Engineering

A small number of related studies make general efforts at using OSINT to find social engineering vulnerabilities.

Huber et al. [2] make use of an organisation's Facebook presence to automatically identify and target its employees. Their tool gathers public information on members from Facebook, then attempts to expand that information through mechanisms like friend requests. Theoretically, their tool then uses Facebook chat to act as a chat-bot, building a rapport before executing a predefined attack (e.g., sending a link). Their evaluation shows that this scheme is impractical due to the overhead associated with imitating a human conversational partner.

Our approach also involves collecting information on employees using the organisation's social media footprint,

[1] Available from: `https://github.com/Betawolf/social-vuln-scanner`

though our search is across multiple online services rather than one, including the business-oriented LinkedIn and the up-to-the-second Twitter as well as Facebook and Google+. Rather than the invasive chat procedure, our system expands its profile of targets passively, through wider searches for a target's online presence.

Ball et al. [3] detail how open source information can be used to construct spear-phishing attacks on an organisation's employees. They manually mine employee information from an organisation's website and gather additional information using the Maltego toolkit, before then using the Simple Phishing Toolkit to create phishing emails based on each employee's interests.

The approach of Ball et al. demonstrates the value of OSINT in this domain, but their method still relies on significant manual workload, whereas we focus on methods which can be deployed as part of a completely automated scanner.

Scheelen et al. [5] attempt to map out a company's structure from online sources, including gathering information for social engineering. In their method, they first connect to the company on LinkedIn and then crawl LinkedIn for a list of employees, then search Facebook for those employees, matching on name, profile picture and location. They prune multiple matches by sending friend requests from 'zombie' profiles which are designed so as to look relevant to the targeted organisation. Their organisational mapping is based on heuristic processing of self-reported roles in LinkedIn profiles.

In contrast to the connections and friend requests utilised by Scheelen et al., our interaction with the target organisation is entirely passive, leaving the target organisation unaware of this stage of information-gathering. While we also resolve the identities of employees, we do this through a more flexible process using a larger and richer set of potential features, as described in Section 4.2.2.

2.2. Identity Resolution

As our design relies on resolving identities across different data sources, existing literature on identity resolution is quite relevant. A variety of methods have been applied in both matching online identities from different social networks and in searching for personal profiles given an existing profile of the same person, examining a range of features.

Obvious features often work well: Perito et al. [6] focused on the identifiability of usernames. As well as contributing a Markov modelling approach for estimating uniqueness of usernames which suggested that they are on average highly unique identifiers, they build and evaluate a classifier which links profiles based on username pairings, achieving good classification accuracy, and suggesting that usernames are an ideal feature for connecting profiles.

Combining features can also prove effective. Irani et al. [7] suggest that a record-matching approach to the problem can be fruitful, with identifiers like last name, birth year and country unlikely to change across records. Working with a wider range of features, Malhotra et al. [8] design an ensemble classifier, with subclassifiers relying on individual features such as profile pictures and usernames.

Social media profiles allow for additional information to be exploited. Goga et al. [9] exploit 'innocuous' social media profile information such as time-stamps, geographical location and writing styles to match user profiles, demonstrating that even where usernames and other traditional identifiers are disguised, users can still be identified based on their usage of the media. Our own method follows from this general design, using multiple subclassifiers on pictures, usernames, writing style, social graphs, content and location.

In contrast to the above studies, where classifiers are trained to connect profiles between two specific networks, our focus is on a system which resolves identities between multiple networks, such that connections can be drawn between profiles on all four of Google+, LinkedIn, Twitter and Facebook.

2.3. Social Engineering Vulnerabilities

Our method relies on linking profiles, a practice which specifically ties to certain vulnerabilities. Linked profiles can be particularly vulnerable to certain social engineering attacks. Chen et al. [10] detail some of these vulnerabilities, and demonstrate that additional details such as phone numbers can be better retrieved when multiple profiles of the target can be linked.

As a complement to this, the absence of a profile on a certain social media network can be a vulnerability in itself. Kontaxis et al. [11] describes the profile cloning attack which lets social engineers use existing information on one person to imitate them on a service on which they do not have an account, along with a detection strategy for this.

More generally, there is a wide body of literature regarding specific social engineering vulnerabilities. We focus our attention on the most pressing social engineering channels which rely on information available online. Krombholz et al. [12] provides a survey of such techniques which we take as instructive in this regard. The general identity resolution approach of pairwise comparison is supported even more broadly by similar approaches in other domains of security analytics [13, 14].

3. Social Engineering Penetration Testing

In order to better understand how OSINT is actually used in real-world social engineering attacks we sought advice from professional social engineering penetration testers. There are only a small number of penetration testers who specialise in social engineering, and most are based at large penetration testing companies. Assisted by CREST[2],

[2] "CREST provides organisations wishing to buy penetration testing services with confidence that the work will be carried out by qual-

interview participants from member companies were solicited for research into the use of OSINT in social engineering engagements. Six professionally qualified penetration testing experts, with knowledge and experience in social engineering engagements, volunteered to be interviewed. Each interviewee held internationally recognised certification in the domain area, reaching a minimum level of 'CREST CRT'[3], allowing level of expertise to be compared and verified to international standards. Each interview lasted 1 hour and 30 minutes.

Participants were asked to discuss their experience of social engineering **attack methodologies**. At this stage of the study, our aim was to determine the real-world attack vectors used by social engineers, and understand their practicality for deployment. This allows us to identify which techniques are more often used, and those that would be preferred. Following this, experts were questioned on the importance of **OSINT data** items for each attack vector, and how much its collection was automated. This included identifying the essential data needed to bootstrap an attack or payload, and the non-essential OSINT data which can still contribute to the effectiveness of an attack. Finally, the experts were asked for **mitigation techniques** for the attack vectors discussed.

A key goal of the interviews was to determine the attack vectors and OSINT data that are used in the real world, and filter out the outliers that are rarely deployed effectively or are embellished in literature as to their effectiveness. We sought to identify techniques that are infrequently used at present, but would see more use in real world attacks, should current tools be enhanced to remove difficulties, such as automated passive collection of data, as suggested in this paper.

3.1. Attack methodologies

The expert social engineers consistently identified the following as the attack vectors used in real-world engagements:

- **Email**: phishing / spear-phishing emails that were used to manipulate a target into visiting a malicious website, or opening a malicious file.
- **Telephone**: voice phishing or 'vishing', used to extract information directly or persuade a target into interacting with a malicious website or previously delivered file.
- **Physical**: gaining physical access to an organisation's site or systems, through use of a deceptive pretext, or delivery of physical media (e.g. drop of a USB stick).

In addition to these attack vectors, our six experts were questioned about the use of online attacks, such as waterholing (strategic compromise of a website known to be frequented by target individuals), and the use of social network sites as an attack vector. Such attack vectors were considered by most to be out of bounds in a contract penetration test, due to reliance on services external to the customer, risk of collateral damage, and invasion of employee privacy. It was noted by the experts that such concerns were not considered by criminals.

Experts were asked to evaluate each individual attack method against the following criteria:

- **Frequency of use**: how often different attacks and deceptions are used in real-world engagements.
- **Effectiveness**: rate of success and detection.
- **Efficiency**: time requirement and level of automation.

Responses were largely consistent in the frequency of use and effectiveness of attack methods used, in terms of rates of success and detection of these attacks. However, it was clear from discussions that success was often interpreted as an overall objective of a penetration test, rather than an individual attack; e.g., from 100 phishing emails sent, 10 may be opened, but 1 may result in a successful compromise of the organisation.

For email-based attacks, all interviewees stated that these were frequently used (more so than any other attack vector), and all but one claimed the method to be successful in the majority of cases, with low detection rates. The level of automation and time frame varied, ranging from almost completely manual to almost fully automated, and from a few hours in one afternoon, to waiting weeks for a response. This reflects the wide range of engagements social engineering penetration testers are involved in.

For telephone-based attacks, 3 interviewees often employed this as an attack vector, 2 did around half of the time, and 1 not all. It was agreed that this was a successful method the majority of the time, with at least one set of credentials (or some other target information) gained in most engagements. Detection rates reported varied dramatically, again depending on the exact nature of the attack and the information sought. This was always done entirely manually, with each call normally lasting 10-15 minutes (except for one interviewee using much shorter phone calls of less than a minute).

Physical access attacks were part of less engagements according to our experts, but still used quite commonly (over 80% of engagements) in most cases (4/6). Success rates ranged from 50% to above 90%, with detection rates reported as being low (except for one report of USB key drops). This was always done entirely manually, with engagements taking at least a day, and sometimes up to a week.

ified individuals with up to date knowledge, skill and competence of the latest vulnerabilities and techniques used by real attackers" (`http://www.crest-approved.org/about-crest/`)

[3] "Analysis and recommendations for standardization in penetration testing and vulnerability assessment" (`http://shop.bsigroup.com/upload/271543/Pen\%20Test\%20Standards\%20Report.pdf`)

3.2. OSINT usage

The main focus of the interviews was the use of OSINT data for the attacks discussed. Following discussion of each attack vector, experts were asked to detail OSINT items that facilitate it, highlighting whether they are essential to the attack process (i.e. an attack cannot occur without this OSINT item), or non-essential. For non-essential items, experts were asked to discuss the degree to which each item contributed to, or accentuated, the success of an attack. Where possible, experts were asked to rate their perceived importance of non-essentail items, so as to provide a point of reference relative to the contribution of other pieces of OSINT data.

Through these discussions, OSINT data items were preliminarily separated into two key information types:

- **Bootstrap**: data which facilitates the attack, usually by allowing targeting of an individual or group of individuals. Consistently, experts reported that whilst target selection focused on those individuals who might be most susceptible to it, the focus was mainly to exclude individuals likely to be less vulnerable, such as IT or security personnel.

- **Accentuator**: data items which are used to enhance the effectiveness of an attack, by adding real-world context to the ploys; such as impersonation of a contact, or inclusion of an event or activity known to be of interest to the target. Accentuator items were rated, (high, medium, or low) based on the the discussions and benchmarked by the ratings given by experts, where appropriate.

In addition to the perceived importance of each item, experts were asked to discuss the process of obtaining the OSINT data items, focusing on time required and level of automation of the process. In this manner we gained an understanding of the resources required to extract each OSINT data item. To understand the rank of importance of OSINT data items, perceived importance to the attack process was compared to the resources required to extract the data, in terms of time and level of automation. In this manner, OSINT data that is easy to obtain (i.e. fast and automated) was ranked more highly, than on requiring increased resources to extract. Furthermore, we are able to identify the level at which OSINT data contributes to individual or multiple attack vectors; flagging those items which bootstrap multiple attack vectors as more useful to an attacker.

The various OSINT items identified by our experts and their nature (as bootstrap or accentuator) are shown in Table 1. Bootstrap (B) items are shown in red, whilst Accentuators (A) are shown in yellow.

In terms of automation of the collection of OSINT data, it was clear from the interviews that whilst there was consistent use of some tools between organisations, allowing automation of basic tasks, the penetration testers tool chains varied greatly between companies, often relying on a pool of scripts, produced in house between penetration testers. Even in cases where OSINT extraction tasks were highly automated, a high proportion of time spent by experts, was focused on verification of the automatically extracted information, prior to its use in an attack.

Table 1: Level of contribution of OSINT data to attack impact. B = Required to bootstrap an attack; A = accentuates an attack.

Item	**Email**	**Phone**	**Onsite**
Name	B		
Name of person with job title	B	B	B
Identity and position in company structure	B	B	B
Name linked to employment by company	B		
Name of new employees	A	A	
Format of email (e.g. inital.lastname@company.com	B		
Specific Email address	B		
Group / generic Telephone number		B	
Direct phone number with name		B	
Cell phone number		A	
Social media posts	A	A	
Social media connections / friends	A		
Social media photos	A		A
Social media hobbies / sports / groups	A		
Email footer / communications sample	A		
Company supplier / partner information	A	A	B
Employee availability / daily routine	A		B
Absence indicators (e.g. out of office reply, Facebook)			B
Files shared on corporate website (PDF, XLS, DB etc)	B		
Identity of facilities manager			B

To put these results into context, we present an example scenario, illustrating how bootstrap data is used to initiate the attack, and accentuator items increase effectiveness.

Example spear phishing email attack:

- An attacker wishes to deliver a malicious spreadsheet to a member of staff within a company.

- HR and finance departments are chosen, due to the

frequency of working with this sort of file and minimal presence of IT personnel.

- The attacker does not have direct email addresses of staff members in the HR or Finance teams.
- Using an automated tool, the attacker locates **the name** (B) of one **member of HR staff** (B) from the company website, and **one name** (B) from LinkedIn.
- Armed with this information and examples of other email addresses from the company website, the attacker is able to deduce the **format for the suspected specific email address** (B) of the two members of HR staff (e.g. firstinitial.lastname@company.com).
- The attacker emails the two employees, with a non-malicious email, requesting legitimate information, to verify the accuracy of the email addresses.
- Members of HR reply, verifying their email addresses and providing the attacker with their **email signature** (A).
- Attacker sends a simple email to one target, containing a malicious spreadsheet disguised as a HR recruitment plan. The sender address of the email is spoofed to be the other member of HR staff located by the attacker, and included in the body of the email is the legitimate email signature of the member of staff; both the familiar sender and their email footer supporting the validity of the message.

During the high-profile social engineering attack on RSA, a malicious Excel spreadsheet was sent to four members of staff within the HR department, along with a simple request stating "I forward this file to you for review. Please open this file and view it". Despite the email being flagged as junk, this was sufficient enough for the recipient to open the attachment and activate the remote access trojan hidden within the malicious file, compromising the corporate network.

3.3. Mitigation techniques

Also discussed during the interviews were social engineering mitigation strategies suggested to clients of penetration tests during the reporting process. The following were mentioned consistently.

Security Awareness training was recommended consistently throughout our interviews as the key mitigation strategy for all the attack vectors used in real world scenarios. It should be provided by default by any organisation, and renewed regularly. Training can help employees understand why a culture of security is important, highlighting the importance of monitoring the information that is available about them online. However, interviewees were consistently dismissive of the effectiveness of the online security awareness training courses that are often the first port of call for companies, normally due to economic restrictions. Instead, they advocated practical training of small groups of staff on real-world scenarios, tailored to a particular organisation. Our participants highlighted that pragmatic training given an organisational context familiar to the trainees, helped to avoid disengagement with training more common to generic approaches, and helped support growth of a security culture within the organisation. To increase familiarity of the training, it was suggested that it should follow a social engineering penetration test, and incorporate the attacks used on the company thus ensuring relatable scenarios and maximising trainee engagement.

Revised security policies and practices: It is important to promote a culture of security in the workplace, policies and best practice guidance aid this. Social media usage policies may indicate what information about their work employees are allowed to post online, or what privacy and security settings should be in place. Security procedures during telephone calls, e.g. challenge and response for IT support calls requiring disclosure of information. This may then be disseminated, and reinforced by annual security awareness training.

Network restrictions: An organisation may want to consider blocking access to certain websites on corporate systems, e.g. social network sites. Many or the interviewed industry professionals also highlighted whitelisting sites as mitigation techniques that are often recommended to organisations, due to the frequent use of typo-squatted URLs in phishing / vishing attacks. It was noted however, that most organisations did not follow the recommendation of whitelisting, due to issues of practicality. Blacklisting of known malicious websites was suggested by all participants, with several also recommending the use of an automated domain name monitoring service, to alert to new registrations of typo-squatted URLs that may indicate an attack.

Company website review: The information on a company website can be a treasure-trove for attackers. Potentially useful information should be reviewed to assess the balance between the need for the information being present, and whether it poses a potential risk; e.g., whether it is necessary to have direct contact information for all employees publicly available. Unfortunately, as identified, much of the information useful to an attacker is that which is required to be disclosed as part of normal business operations.

Further Social engineering penetration tests: Annual social engineering penetration tests, followed by an ongoing programme of staff education was suggested by all participants as a key method of identifying vulnerabilities, and evaluation of the effectiveness of other mitigation techniques over time. Experts commented that the scope of penetration tests was often limited in terms of access to personal employee information, for reasons of ethics, blocking access to some of the most useful OSINT data items shown in Table 1. Due to economic restrictions,

many penetration testers highlighted that the duration of penetration tests was also often highly restricted, leaving a minimal time for collection of OSINT data, that may not be representative of a determined persistent adversary. In these cases, white-box penetration testing was suggested, so as to test the underlaying security procedures in the most time efficient manner.

3.4. Summary

From the data gathered from these interviews we are able to rank potential real-world attacks by effectiveness and efficiency. Armed with knowledge of which OSINT data is required to facilitate these attacks, which information contributes to their effectiveness, and an understanding of the ease with which OSINT is obtained by an attacker, we can better understand which pieces of OSINT data lead to vulnerabilities that can be exploited. In the following sections we show that these OSINT items can be found automatically to reveal an organisation's *social engineering attack surface.*

4. Automatically Identifying Employee Information

To automatically assess the vulnerability of an organisation's online presence with respect to the risky information items listed in Table 1, two critical challenges must be overcome. Firstly, the accounts of employees must be identified from within the online footprint of the organisation, as employees are the critical targets on which information is gathered. Secondly, extended information needs to be retrieved on these employees. Both challenges are addressed in this section.

4.1. Distinguishing Employees from Social Media Followers

Employees are likely to be a minority amongst the followers of an organisation in online social networks. Business partners, customers and competitors also interact with organisations on social media, and effort invested into gathering information on these accounts will be largely wasted for the target-oriented social engineer, and by extension it will be wasted for an automated system approximating them for the purpose of a penetration test. The problem then becomes as follows:

Given a profile of an organisation O, its set of employees O_E and set of social media profiles of affiliates O_A who are linked to O through its friends list or public interaction on social media, how can we select from O_A those profiles which belong in O_E?

4.1.1. Data Source

In attempting to evaluate performance of any method, we need some source of ground truth data. Certain quarters of the business world have helpful practices when it comes to revealing employee information that can be helpful in this regard. The Law Society indexes legal firms, with links to the websites of a large number of these. From an organisation's homepage, we can often automatically derive two further values:

- the official Twitter account for the organisation;
- a link to a 'roster page' which lists the names, and sometimes positions, of employees of the organisation.

From the Twitter account for the organisation we can extract the list of accounts for followers and followees — the affiliates O_A of the organisation. These profiles can be downloaded via the Twitter API to be inspected for features indicating they belong to O_E.

From the roster pages we can extract the names of employees (using the Stanford NER tool [15]). This gives us a list of known employee names O_E.

A dataset consisting of 17 companies, 3,753 affiliate profiles and 448 employee names confirmed from roster pages was extracted using this method, with an average of 221 affiliates per company. The companies selected were drawn from the first 500 companies in the Law Society's index, filtering only those which matched our requirements. For each affiliate, an automatic string comparison method was used to compare their name to the appropriate roster list, allowing complete or partial matches based on name components. Where this automatic comparison found a match, the profile was flagged for manual review to check that the individual was the same person. After this stage, each affiliate is coded as either matching or not matching the roster page. Only 20 such matches were confirmed.

Features corresponding to those described below were also calculated for each employer/affiliate pairing. For evaluation purposes, the feature regarding whether an organisation's website refers to the name of a profile specifically excluded the roster page being used to generate the ground truth.

4.1.2. Features

As we presume the external social engineer's viewpoint, we consider only features of the organisation and the affiliate's online footprint. There are a number of such features which would be indicative of an employment relationship:

1. *FollowedBy/Following*

   The specifics of the identified connection between the organisation and the affiliate may be important in distinguishing employees from non-employees. It is plausible, in networks like Twitter where connections are unidirectional, that a popular organisation may have many followers, but would be more selective in the users which it follows. This distinction may also work in reverse: bidirectional connections may also distinguish employees from e.g. celebrities the company is interested in, who would not be expected to reciprocate the interest.

2. *OnWeb*

   The organisation's public website is, as ever, a rich source of information for a social engineer. Many organisations host a roster page on their website which reveals some or all of their employees to the general public. Even amongst organisations which do not host such a page, material such as press releases and self-promotion text often incidentally refers to employees by name. These names can be considered highly identifiable for individuals connected to the organisation, though there is also some reason to expect Type I errors due to e.g., partner organisations and prominent customers also being incidentally mentioned.

3. *HasFirmName*

   Another revelatory area would be the identification of the company in the profile text of the affiliate. Many professionals will state their employment status as a key part of their identity (e.g. "Marketing Director at EXAMCORP"). However, this behaviour is not necessarily uniformly adopted, and even where it is adopted it can be challenging to automatically recognise the information being sought; an organisation's name can be expressed in many forms, some of which may be unintuitive abbreviations or acronyms.

4. *MentionsEmployee/MentionsEmployer*

   Social media references may be indicative of a close connection between profiles. This can take two forms: the organisation referencing the employee's handle or name in its online posts (congratulation messages, for example) or else the symmetrical case of the employee referencing the employer's name in their online posts. The issues with the previous two items are both relevant here: employers may talk about competitors, partners or customers as well as employees, and employees may refer to their employer by an unintuitive name or not at all.

5. *FollowedMatches/FollowingMatches*

   The network topology of nodes can also be an indication of closeness. Employees are likely to move in the same circles as their employer organisation, and as such their social media contacts lists may overlap more than those of customers or competitors. This common topology could be an especially useful identifier where name-based methods are insufficient.

4.1.3. Method

Our aim is to build a classification system which separates employee from non-employee affiliates. We desire a solution which can be readily implemented in an automated scanning tool and the operation of which can be interpreted by a social engineer to provide feedback. As such, we chose to apply a decision tree classifier, using the features outlined above.

A C5.0 decision tree from the C50 R implementation [16] was evaluated on a 90:10 training/test split of the dataset using a cost matrix which penalised false positives.

4.1.4. Results

The decision tree from a training iteration is reproduced below.

```
OnWeb in {MAYBE,NO}: NO (3331/2)
OnWeb = YES:
:..HasFirmName = YES: YES (7/2)
   HasFirmName = NO:
   :..FollowedMatches > 2: NO (7)
      FollowedMatches <= 2:
      :..FollowingMatches <= 1: YES (23/13)
         FollowingMatches > 1:
          :..FollowedMatches <= 1: NO (7)
             FollowedMatches > 1: YES (2/1)
```

The most valuable feature was `OnWeb` – whether the affiliate's name appears on the website outside of the roster page, a negative or only partial match result here excluding a majority of non-employee affiliates. The following features were whether the employee mentions the organisation's name in their profile text (`HasFirmName`) and then the topological data: the number of users which both profiles are followed by and following (`FollowedMatches`, `FollowingMatches`). Of interest here is that large numbers of followed-by matches are actually a negative predictor of employment – this may reflect a pattern whereby competitor companies are followed by the same userbase, so users which have the same followers as an organisation are more likely competitors than employees.

A stratified ten-fold cross-validation using different portions of the data for training and test sets produced an average *recall* of 0.9 and *precision* of 0.58, which combine as an overall *f1-score* of 0.65 [4]. For the purposes of calculation, a true positive is defined as correctly identifying an employee, and a true negative would be correctly classifying a profile as a non-employee according to the roster page information. Figure 1 plots precision and recall for individual trials. Given the low base rate of confirmed employees (20 out of 3,753), this would appear to be acceptable performance. Although the mediocre precision of the method remains concerning, this may be in part explained by real new or ex-employees who are not listed on the roster, or individuals which could be functionally as critically involved with the company as an employee. The coverage of known employees is effective across the evaluation.

[4] Precision and recall are preferred to accuracy in cases with imbalanced classes, as here, because otherwise it is easy to produce misleadingly positive results by e.g. always predicting the majority class (most profiles are not employees).

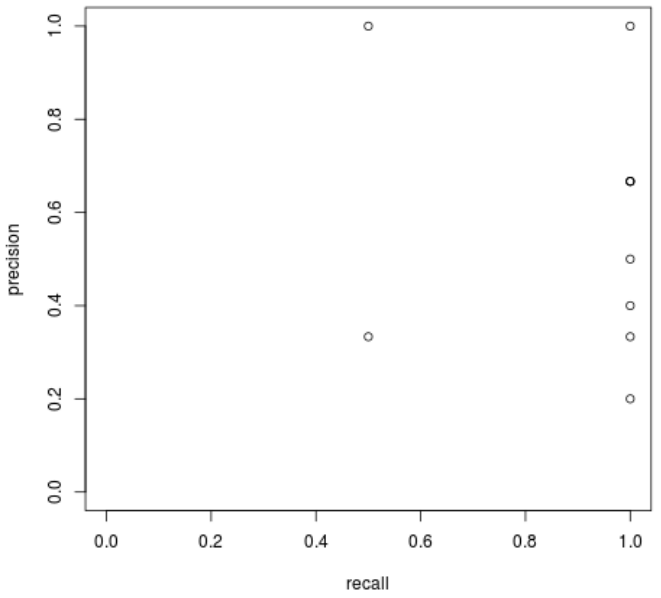

Figure 1: Individual cross-validation results for precision and recall.

4.2. Probabilistic Identity Resolution across Multiple OSNs

Once an employee's profile has been identified, a social engineer will benefit from gathering as much of their online footprint as possible, gathering more context to use in attacks against them. A means of doing this would be to identify the same individual on another online social network, in which additional data may be available. Employees may, for example, maintain both a professional Twitter account which advertises their connection to their employer and a personal Facebook account which does not. Resolving these two identities allows the social engineer access to data available in the personal account in attacks directed at their professional role within a company.

Previous studies [17, 9] have shown that this process can be automated for re-identification of individuals between certain social networks. Our aim is specifically *to resolve identities across multiple online social networks, rather than between only two networks as is typical in the literature.*

4.2.1. Data Source

Our classifier is trained and evaluated on a challenging realistic data set gathered via the Google+ social network. We targeted our efforts on classifying links to other identities on three of the largest online social networks: Facebook, Twitter and LinkedIn.

The Google+ social network includes an "other profiles" attribute which highlights profiles of the account holder on other online social networks. Using a method adopted from Gonzalez et al [18], we randomly sampled 1,161 Google+ profiles from the network.

For those profiles from this sample which included an "other profiles" attribute, the referenced profile was downloaded to the extent permitted by that social network's API – acting only as an application or developer account, with no effort made at invasive methods such as issuing friend requests. These results form positive examples of matches. To create the appropriate conditions for classification, an automatic process issued a search to each social network for the name of the profile on Google+ from which the link was made – emulating the behaviour of a social engineer who has found only the Google+ profile and is now seeking the same individual on other networks. The profiles resulting from this search were collected as negative examples. This sampling process is designed to avoid biases in the dataset being used for evaluating this component of the system, and is more fully described by Edwards et al. [19].

As is typical in record linkage efforts, classification was attempted within blocks of probable matches. In this case the blocks were defined by the search term used to retrieve the negative examples, combined with the relevant positive examples revealed by the "other profiles" attribute of the Google+ profile. The final dataset consisted of a total of 8,402 comparisons between profiles from Google+ and the three other networks, including 89 positive matches – a base rate of just over 1%.

Within each pair comparison, all of the subclassifiers described below examined the available profile data on the pair of profiles and reported their similarity estimate between [0,1], producing a comparison vector of 8 ratings. Where data was not available for an attribute, the classifier would respond with a similarity of 0 – reflecting a real lack of similarity where the data cannot be observed.

4.2.2. Features

A number of different features may identify the same person on different media, so the approach taken here is an ensemble classifier, which makes its decision about whether two profiles can be matched based on the reported matches of sub-classifiers working on specific profile features, discussed below.

In most cases more advanced methods exist which would produce more accurate comparison results for each feature, but our primary focus has been on employing unsupervised methods which release us from the requirement of obtaining training data for subclassifiers and which are relatively computationally inexpensive.

Each subclassifier reports a confidence in the interval [0,1] that the two profiles presented reflect the same person. These features can then be weighted and combined with a logistic regression model.

1. *Name Subclassifiers*

   Names and usernames are some of the best identifying features, and are near-ubiquitous in online services. As such, two subclassifiers make use of names from user profiles.

a) The first subclassifier compares the proportion of name components from each profile which are exactly matched in another profile. For example, comparing a profile with the name components "John" and "Hancock" and "@JHC" to another with the components "John" and "Smith" and "+JS202" would produce a $0.3\overline{3}$ match due to the shared first name. Conversely, comparing the second profile with a profile with name components "John" and "Smith" and "@JS_work" would produce an overlap of 0.66 due to a full name match.

b) The second subclassifier returns a similarity measure based on the edit distance between the two profiles' most representative names. For this we use the Damerau-Levenshtein edit distance calculation, where the classifier's similarity measure is given as follows:

$$1 - \frac{edit\_distance}{longest\_string\_length}$$

Continuing the example from above, the representative names "John Hancock" and "John Smith" would produce a similarity measure of $0.41\overline{6}$ — an edit distance of 7 operations over the 12 characters in the longest string. Conversely, comparison of the latter to "John J. Smith" would produce a similarity measure of $0.8\overline{33}$, with a distance of 2 operations over the 12 characters in the longest string.

2. *Profile Picture Subclassifier*

   Avatars are images selected by a user to convey their identity, and as such there is some reason to believe that users will make use of the same image on different networks, and as such it can be used to link their profiles.

   The primary avatar image from each profile is resized to comparable thumbnail dimensions, and then the euclidean distance between the histogram of each thumbnail is computed. This provides a simple visual comparison of the two images. Although highly sensitive to translation, rotation and other simple manipulations, the aim of this comparison is to determine if the avatar images are essentially identical.

3. *Activity Time Subclassifier*

   The times at which a user updates their profile provide behavioural clues about both their geographical location and their habits and routine, both of which are useful in identification.

   We base our time comparison of profiles on the work of Atig *et al.* [20]. Timestamps from updates to social media are sorted into one of six four-hour bins which make up an activity profile. Each slot in a profile is compared to a higher (20%) and lower (8%) threshold of activity. Those slots with more than the upper threshold's proportion of activity are considered 'high activity' slots for the profile, and those with less than the lower threshold are considered 'low activity' slots. If two profiles share a high or low activity slot, this increases the rating of similarity between the two profiles.

4. *Writing Fingerprint Subclassifier*

   Writing style similarity can be highly useful in identifying individuals, and the authorship analysis literature contains many well-performing supervised solutions (e.g. Afroz et al. [21]).

   We analyse the textual content of posts made by each profile by counting the usage of a number of function words. Function words are words such as 'it','some','if' and 'there' which have little lexical meaning, but form the structure of sentences. They can be highly indicative of writing style [22]. Our method counts the proportion of text from both profiles which is composed of any of a list of 70 common function words, and calculates the euclidean distance between these proportions to quantify the similarity in writing style.

5. *Link Analysis Subclassifier*

   As well as writing style, matched content can reveal a similarity or connection between individual profiles. Hyperlinks are unique identifiers of content of interest in online documents. People sharing the same link are likely to have similar interests or read the same news source, and may in fact be the same person promoting a link of interest to different social groups on their multiple social networking platforms.

   We count the proportional overlap in the set of hyperlinks found in the user-generated text associated with a profile. This subclassifier thus captures the behaviour of users promoting links of interest on multiple platforms.

6. *Friends Subclassifier*

   A person's social graph in one social network is likely to resemble their social graph in another. We make use of this to compare profiles by matching the names of friends from both profiles. Wherever a friend's name in one profile matches a friend's name in another (according to the edit distance method of subclassifier 1(a), and a threshold of 0.8 similarity), the similarity is incremented, up to a maximum which would represent all friends in one profile having a strong name-based match in the other profile.

7. *Geographic Subclassifier*

   Geographic location can be highly identifying, when the data is made available in sufficient resolution. We create a list of location pairs as the product of the known locations associated with two profiles, and

evaluate whether each pair is 'near' the other, where 'near' is defined as the haversine distance between two long/lat points being below a threshold of 10 kilometers. The total number of 'near' pairs determines the geographic similarity of the two profiles.

4.2.3. Method

The dataset of comparison vectors was divided into a stratified 90:10 training/test set. A binomial logistic regression model was fitted to the training data, including terms for interactions between each of the subclassifier outputs, as we would expect the subclassifiers to generally support each other.

4.2.4. Results

Table 2: precision, recall and f1-scores for classifier thresholds between 0 and 0.9

threshold	precision	recall	f1-score
0.0	0.01	1.0	0.02
0.1	0.25	0.625	0.357
0.2	0.44	0.5	0.471
0.3	*0.8*	*0.5*	*0.615*
0.4	0.8	0.5	0.615
0.5	0.8	0.5	0.615
0.6	0.8	0.5	0.615
0.7	0.8	0.5	0.615
0.8	0.8	0.5	0.615
0.9	0.75	0.375	0.5

Threshold values were applied to the model's output probabilities to enable classification and produce *precision* and *recall* measures, and the combined *f1-score*. A range of these values for a sample iteration is presented in Table 2. For the purposes of calculation, a true positive is defined as correctly identifying a matched profile, and a true negative is correctly identifying non-matched profiles. The italicised row 5 highlights the greatest *f1-score* of 0.615 at the threshold of 0.3, which represents the optimum threshold with equal balance being given to precision and recall. In this iteration, as in others, precision remains stable for a range of threshold values past 0.3, suggesting that low model output probabilities are efficient at capturing most of the non-matched accounts. The inclusion of a threshold value of 0 highlights the difference between model performance and the base rate for the data.

Taking this threshold, a stratified ten-fold cross-validation found an average *precision* of 0.64 and *recall* of 0.46, for an average *f1-score* of 0.51. The AUROC for the model remained high through all iterations, with an average AUROC of 0.96. The ROC plot in Figure 2 compares the ROC curves from each iteration of the cross-validation. The high AUROC figure indicates that the model has good discriminative power, and that real identity matches score much better on the model output than mappings to other

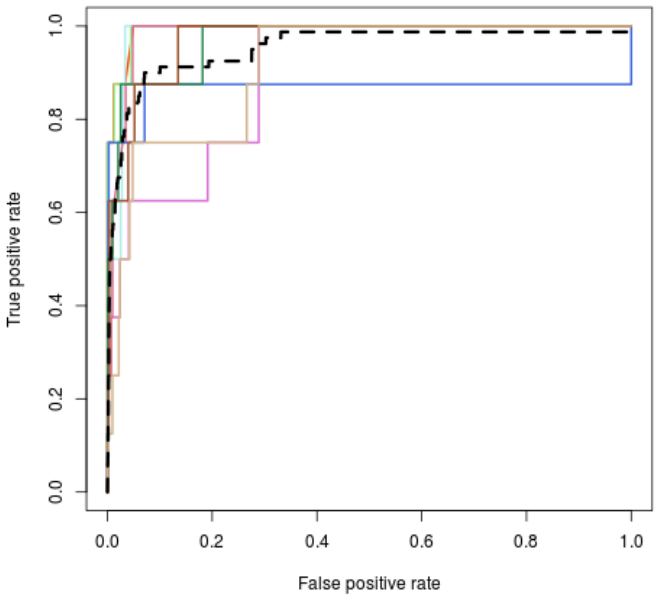

Figure 2: ROC plots for model validation

affiliates. In comparison, Figure 3 plots precision and recall over cross-validation runs. The mean AUPRC is 0.48.

Regarding features, the majority of the predictive power (as derived by the regression model) was carried by the name, time activity profile and avatar comparisons, along with their interactions, but the link activity subclassifier was also highly predictive. The network-based comparisons suffered from high standard errors, while our stylometry and geography comparisons appear to have been too often lacking sufficient data for comparison.

5. Application to Critical Infrastructure Organisations

Combining the automatic solution for identifying an organisation's employees with an identity resolution system which can be applied across social networks allows for the collection of rich data about the organisation's employees. For a human attacker, manually tracing the link between an organisation's website and its employees' disparate social media activity is not that difficult, but at the same time it would be burdensome and intrusive for the organisation to replicate this process as part of a recurring vulnerability assessment.

Our automated methods allow for a minimally invasive social engineering vulnerability scan to be launched against an organisation's web presence. Starting with nothing but the organisation's homepage URL, it is possible to locate employee accounts and from there determine the availability of names, photographs and activity information which would greatly aid a social engineer. This in-

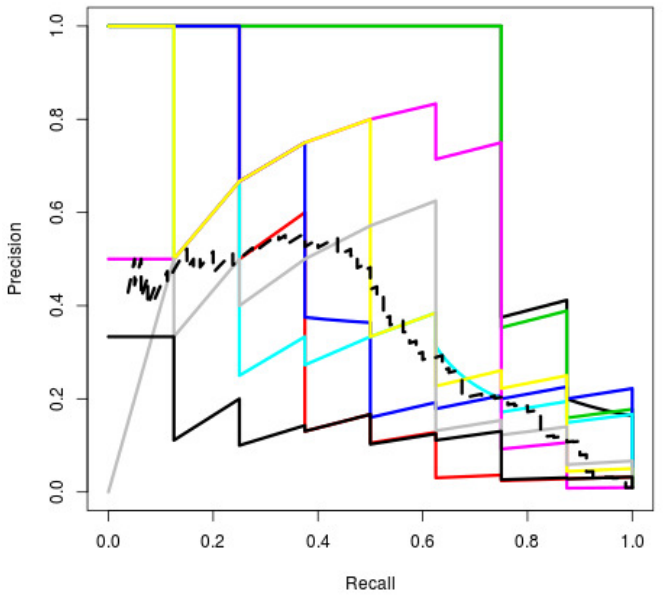

Figure 3: PR plots for model validation

formation can then be used as an appraisal of the organisation's vulnerability to social engineering, to complement the more detailed feedback a human assessor might provide with a less intrusive measure of the effectiveness of new policies and initiatives.

As a demonstration of the vulnerability which may be exposed by the social media footprint of employers and employees, we ran an automated scan on the digital footprint of multiple real critical infrastructure organisations: specifically, water, gas and electricity companies. This selection was based on the increasing focus such organisations have received over recent years, particularly with regards to their use of industrial control systems (ICS), to monitor, control, and automate physical operational processes. To date this focus has largely been attributed to technical shortcomings, demonstrated in the sharp rise of disclosed vulnerabilities [23], and neglecting the importance of social and organisational factors. An initial step towards understanding the potential impact of social engineering on ICS was discussed by Green et al. [24], however this work focuses solely on malicious emails, with vulnerability assessments achieved through the use of interview data.

Using a fully automated procedure based on the systems we have evaluated, we assessed the online footprint of these companies, firstly to identify employees from amongst their social media connections and secondly to gather those information items we have identified (with reference to Table 1) as important to social engineers targeting the company via these employees.

The input in each case was the URL for the target company's website. The website was crawled and the content scraped – amongst the information extracted were contact details such as email addresses and phone numbers, but most important were URLs directed at the company's social media presence. These URLs were detected and resolved to the indicated account, information on which was then retrieved via the appropriate social network API – where an account could not be found, the run was terminated. From these accounts, lists of affiliated profiles were collected, and our decision-tree classifier identified employees from amongst this set. Searches were issued for the employee profile display names on each of the four networks, and the trained logistic regression model was used to identify any likely matches to the indicated employee profiles, joining this information to the stored profile on that employee. The output for each organisation was a summary of the number of important social engineering information items available. For employee items, each category of information was only counted once per person (i.e. recovering multiple names or status posts does not inflate the count).

As broken down in Table 3, our employee identification method filtered the total of 15,551 affiliate profiles down to 128 employee profiles. With regards to the bootstrap and accentuator totals listed:

- employee names were extracted from both the website and connected social media presences;
- email addresses and phone numbers were extracted from the organisational website using pattern-matching (checks were also made for additional contact data from social media, but none was found);
- whether activity information was available was judged based on whether multiple posts per day were available for a significant proportion of each employee's observed timeline;
- documentation was counted via the presence of PDF files in the organisational website;
- information on friends, textual updates and photos were extracted from social media profiles.

Alongside bootstrap data specific to the company's own online footprint, for each of these employees, enough bootstrap-threshold data was automatically extracted from their online profiles to enable a targeted attack, and multiple accentuating information items were available. The bootstrap figures reported in Table 3 separate those instances where a bootstrap information item was available on an employee profile (E) from the bootstrap information items which were extracted directly from the organisation's website (O). In all cases, the organisational website provided a greater volume of bootstrap data, although the categories were for the most part non-overlapping, with *activity* information only available via social media. The presence

Table 3: Outcomes for target companies. E is total bootstrap information extracted from employees social media profiles, O is total bootstrap information extracted from the organisation website.

		targets		bootstrap							accentuator			
		affiliates	employees	name	email	phone	activity	docs	E	O	friends	text	photos	total
Water	1	1792	29	29	61	148	25	1431	54	210	26	26	29	81
	2	502	2	2	9	85	0	692	2	95	0	0	2	2
	3	3258	17	17	38	76	12	1465	29	115	14	13	17	44
	4	4483	44	44	45	196	30	90	74	321	33	32	44	109
	5	508	3	3	14	41	2	683	5	56	3	3	3	9
Gas	1	500	1	1	27	58	1	629	2	86	1	1	1	3
	2	930	11	11	14	31	7	341	18	46	9	7	11	27
	3	471	5	5	18	26	3	352	8	45	4	3	5	12
	4	463	3	3	25	74	3	197	6	100	2	3	3	8
Power	1	1815	12	12	52	78	11	16	23	131	12	12	12	36
	2	454	0	-	0	3	-	0	-	3	-	-	-	-
	3	475	1	1	10	6	1	639	2	17	1	1	1	3
	4	1374	10	10	37	79	8	27	18	117	9	8	10	27

of documentation rather than its volume counted towards the organisational total for bootstrap items.

Organisational footprints contained the most direct contact information – all phone numbers and email addresses retrieved were automatically mined from organisational websites. Employee social media profiles provided a complement to this, turning up large amounts of visual, text, activity and social information for most individual employees.

The information items revealed by the employee profiles of these companies were best-suited to email-based attacks. While email addresses themselves were only found on organisation websites, social media profiles combined necessary attack information – the employee's name – with their identification as an employee of the company, and other information which would allow social engineers to craft a phishing attack to the employee's personal interests (e.g. a link purporting to be for a competition to win something the employee has an interest in) or social circle (e.g. imitating a coworker or personal friend). It is worth noting that many email-based attacks could also be delivered via social media communications.

Phone-based attacks were also given strong support – the combination of a phone number retrieved from the company's website and knowledge of the names of employees and other context allows social engineers to talk their targets into compromising organisational security. The lack of direct phone numbers to contact employees is the only hindrance to such attacks.

On-site attacks are the least well provided for by automatic examination of public OSINT items, with challenges of inferring a company's supplier or partner organisations hindering some attack schemes. Social media does however often provide the activity information about employees which reveals their work routine, allowing social engineers to identify, for example, periods when employees would be out to lunch and buildings could be more easily accessed. The availability of geo-location data is also significant for these attacks, as recent work demonstrates [25].

6. Automated scanning as a mitigation strategy

Prevention of security breaches resulting from social engineering attacks is notoriously difficult, with human error or human manipulation playing a large factor in the majority of high-profile cases. As shown, OSINT provides a key component to, and assists, many social engineering attacks. Currently, for an organisation to assess its own online footprint, including the presence of its employees, and the risk this poses, intensive manual effort is required, and usually the expertise of a social engineering penetration tester. Furthermore, this process must be repeated at regular intervals as information and content is constantly updated. The automated vulnerability assessment tool developed as a product of this research is able to provide an assessment of an organisation's *social engineering attack surface*, as demonstrated in Section 5. We propose that such a tool is useful for organisations to utilise as part of an ongoing assessment of their risk from online footprints and OSINT assisting potential social engineering attacks. To demonstrate the tool as a vulnerability scanner and mitigation strategy, we attached a simple automated reporting mechanism to the tool, which provides information such as:

- Which organisation-level websites and social media profiles were identified.
- The number of people found affiliated to the company, and how many of these were detected as employees.
- A list of the OSINT item types collected and quantities (as in Table 3), along with a high level description of the types of attack these could enable.

- A list of mitigation strategies to consider, prioritised by the list of OSINT items found. These were informed by the expert interviews described in Section 3. These include:
  - Security awareness training.
  - Revised security policies and practices.
  - Network restrictions.
  - Company website review.
  - Social engineering penetration test.

In order to assess the usefulness of the final tool in a real-world environment, we conducted trials with 5 organisations, who we approached for their cooperation. Anonymity was assured for each organisation. The organisations provided URLs to use as input to the tool, and were then in turn presented with a report.

The organisations taking part in the study were:

A. The IT services department (c. 150 employees) of a large organisation (c. 5,000 employees). The head of IT security was interviewed.

B. A small legal firm (c. 50 employees). The Chief Executive Officer was interviewed.

C. A small IT security company (c. 20 employees). The managing director was interviewed.

D. A small-to-medium legal firm (c. 100 employees). The head of IT was interviewed.

E. A small IT company (c. 10 employees). The head of security was interviewed.

Once provided with the report, each organisation was interviewed at length to garner their feedback on the findings, and its usefulness to the organisation. Questions included "Have the results of the scanner affected your views on publicly available information on your company and staff?" and "What do you intend to do now that you have the results?"

In terms of altering perception, there were no immediate indications that the scanner produced results which would shock an organisation into adopting a more strict security culture. However, some companies admitted that the risk factors analysed heightened their awareness of certain risks which they were previously not aware of. This was especially true for those with minimal security knowledge and background.

The tool's output was seen to be of benefit for highlighting potential issues, and useful for supporting the argument for, and implementation of, future mitigation strategies. Furthermore, most companies stated they would like to use the tool on a regular basis to evaluate progress and the effect of mitigation strategies, e.g. Company E stated:

> "We can use this to benchmark ourselves and include this in our monthly staff meetings to check progress on our own cyber security and risk of social engineering. That should help to bring a cyber security culture within the company."

And Company A stated:

> "Once you've run things and delivered training, it's good to use that as a metric to see how successful that was and how much of an impact that's made."

It's important to note that we do not name individuals in our results, thus maintaining anonymity, and only report aggregate figures. Some of the feedback indicated that organisations would like more specific information. For example, Company A stated:

> "Maybe give some examples that are specific to [the organisation]. If I'm looking at [this tool] from a CEO point of view, to be able to identify this [example] to the statistic and say this person said such and such, I think it'll hit home a bit more."

Company A also felt that the outputs of the scanner would be more useful if each department within the organisation could be distinguished. While this still provides a level of anonymity, it can reveal a more detailed insight and could allow management or staff responsible for security to focus on departments which seem high risk:

> "Because not everybody is forward facing so I'd be less bothered about them."

7. Discussion

7.1. Automated social engineering scans vs. penetration testing

It seems clear that an automated social engineering vulnerability scanner, such as that developed for this research, is of use for organisations as an initial, and ongoing, assessment of risk from their online footprints and OSINT. The automated scan is clearly less expensive to run than a manual analysis, with adequately thorough manual scans taking an individual hours to perform. The scan can be run with no human intervention with simple initial variables set. It is also possible for an automated scan to cover many more individuals than a human could reasonably analyse, and thus potentially find more OSINT items of risk.

A social engineering vulnerability scanner could not, and should not, replace a social engineering penetration test, much like an automated vulnerability scanner (e.g., website scanners) could not adequately replace a full penetration test. However, there is scope for both to be used

alongside each other. A vulnerability scanner provides cheap and simple results to perform an initial assessment of an organisation's vulnerability to social engineering. It may provide an indication to what extent a full penetration test involving social engineering is necessary. It also provides important assessment to small- and medium-sized organisation who may not be able to afford a full penetration test, providing pointers to mitigation strategies that any organisation could consider putting in place.

7.2. Reporting granularity

An interesting discussion point is the granularity provided to distinguish individuals in the report. Some companies requested further breakdown into employees from individual departments — this could also be applied to different roles. However, it would be possible with the data collected to go further and produce a list naming individuals who are considered high-risk or who exhibit specific OSINT items. Whilst technically possible, there are ethical and moral issues to consider here. Singling out individuals in this way may highlight targeted mitigation strategies and alert an individual to issues s/he may not be aware of. But it may also cause distress and lead to disciplinary action. It would seem to be particularly unreasonable to single out individuals for risky behaviour when awareness training has not been provided and relevant policies have not been set out prior to the scan taking place. A similar issue is raised during social engineering penetration testing — should an individual who is found to pose a security risk (e.g. gives away a password) be named to management? Views on this seem to be mixed, although anecdotal evidence suggests that human resources departments would raise serious concerns about such practice. It should be noted however that a malicious attacker using social engineering techniques is unlikely to deliberate on the issue, so from a security standpoint the benefit of identifying the precise 'point of failure' can be seen. As an analogy, in a technical penetration test it would not be satisfactory if a report stated that one input field on the company's websites is vulnerable to SQL injection without specifying precisely which input field on which webpage.

7.3. Dual-use concerns

As with any vulnerability scanner, the prototype can be used for harm as well as good, with vulnerabilities highlighted to potential attackers as well the organisation. This is mitigated somewhat by aggregated statistics being presented rather than specific details about vulnerable individuals, as discussed. Also, the tool does not present any information that could not be gleaned through manual analysis by an attacker. Indeed a determined attacker would be likely to find additional useful information and specific individuals to target. One conceivable misuse of the tool would be to automatically scan a large number of organisations, and pick out those particularly vulnerable to social engineering attacks, or particular ploys. The time taken to run each scan, mainly due to rate limiting, would mitigate this approach somewhat. It is hoped that the development and use of the tool can highlight OSINT and social engineering risk to a wider range of organisations who may not have considered such risks adequately before, and also provide training opportunities, mitigation strategies and benchmark assessments of online footprints.

8. Conclusion

Social engineering attacks are a potent threat to organisational security, and open-source intelligence provides vital data which enables attackers to carry them out. In this paper we have used expert guidance to determine which information is of practical value to social engineers, and demonstrated that significant elements of such attacks can be automated in a passive manner. Further, we have demonstrated that this approach is fruitful. Employees can be identified amongst large online crowds and selected for automatic data harvesting, and when they are, large amounts of information valuable to social engineers can be extracted. This extends to information on profiles which the employee does not necessarily realise can be connected to their work identity. As the Internet of Things and other media trends expand the range of personal information which is made available about individuals, organisations must become increasingly aware of the threat vector which leads from seemingly innocuous personal digital habits to compromise of their staff and systems.

While this has implications as a threat to organisational security, this technology can also be used as a tool to harden an organisation's online presence. To reinforce mitigation strategies, we contribute an automated tool that can be used by penetration testers to passively examine the vulnerability of an organisation, which may be used as a means of evaluating the real effectiveness of organisational mitigation strategies such as training events and updated policies.

8.1. Further Work

Social engineering attacks bridge an employee's personal life with their professional role, and as a result vulnerability assessments must likewise do so — managing this without being unduly aggressive is a delicate balance to strike. In this work we have taken the stance that reporting summaries of information which is freely available on public networks would not be overly invasive of employee privacy, but further examination of expectations could well be a profitable area for study. Much may rest on how resulting vulnerabilities are reported – can employee-level vulnerabilities be reported in good conscience to enable targeted interventions?

We have highlighted the use of a passive vulnerability assessment as a mitigation strategy to help organisations to reduce their social engineering attack surface.

This could be combined with more active countermeasures, such as phishing email susceptibility tests as described by Finn and Jakobsson [26], or by creating honeypot social media accounts in a similar manner described by Lee et al. for uncovering social spammers [27]. Furthermore, Kotson et al. [28] and Dewan et al. [29] propose that organisations could use natural language processing techniques to maintain awareness of their online footprint, to be compared with received phishing emails, allowing identification of collection of OSINT by an attacker, and potential early warning of an Advanced Persistent Threat (APT). During the course of our interviews, experts consistently highlighted awareness training as the best way to combat social engineering attacks. Whilst some technical countermeasures were recommended by these experts, solutions focused on detecting the payload of social engineering attacks (e.g. domain whitelisting / blacklisting, domain monitoring for typo squatting), rather than detection of more sophisticated social engineering attacks. Solutions to detect such sophisticated attacks are emerging [30], but still fall short of successfully detecting against sophisticated text-based attacks.

Our technical contributions could be improved. Our approach has been to start from simple and readily deployable methods which can be included in a tool. The precision of our employee distinguishing system may be improved with the introduction of a richer feature set and better training data, and alternative classifiers could well prove more suitable for the challenge than the C5.0 decision-tree. We are already aware that better-performing feature sets are available for cross-platform identity resolution (e.g. [9]), and these would be fruitful areas for improvement.

Our current implementation focuses on application to the dominant Western, English-language social networks, but the majority of the general process generalises well to other social networks and other languages. For other English-language social networks, the tool can be extended through the development of a module following the template used by the existing four modules, which extract profile information into a standard internal representation. When crossing languages, structural information such as hyperlinks and user relationships can be managed by the existing framework, but the writing fingerprint subclassifier in particular will need to be updated to reflect function-word lists for the target languages, and informed about the appropriate word-list to use for a given corpus. The existing modules could also be redesigned around a web-scraping approach, to bypass authorised API limitations, while feeding into the same core process.

Finally, further end-user evaluations of our tool would help tune operation and output to meet the needs of penetration testers. While certain elements of an online footprint require human attention for their value to be revealed, the ability to automatically quantify certain threats could act as a force multiplier for effective security testing, but this requires more attention and development than typically emerges from research tools.

Acknowledgments

The authors would like to thank the EPSRC for its financial support through Lancaster University's Impact Acceleration Account (Grant Reference EP/K50421X/1).

References

[1] T. N. Jagatic, N. A. Johnson, M. Jakobsson, F. Menczer, Social phishing, Communications of the ACM 50 (10) (2007) 94–100.

[2] M. Huber, S. Kowalski, M. Nohlberg, S. Tjoa, Towards automating social engineering using social networking sites, in: International Conference on Computational Science and Engineering, 2009 (CSE'09), Vol. 3, IEEE, 2009, pp. 117–124.

[3] L. D. Ball, G. Ewan, N. J. Coull, Undermining-social engineering using open source intelligence gathering, in: KDIR 2012: Proceedings of the 4th International Conference on Knowledge Discovery and Information Retrieval, Barcelona, Spain, October 4-7, SciTePress-Science and Technology Publications, 2012.

[4] S. Creese, M. Goldsmith, J. R. C. Nurse, E. Phillips, A data-reachability model for elucidating privacy and security risks related to the use of online social networks, in: 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2012, Liverpool, United Kingdom, June 25-27, 2012, 2012, pp. 1124–1131.

[5] Y. Scheelen, D. Wagenaar, M. Smeets, M. Kuczynski, The devil is in the details: Social engineering by means of social media, in: A Project Report on System & Network Engineering, Universiteit van Amsterdam, 2012.

[6] D. Perito, C. Castelluccia, M. A. Kaafar, P. Manils, How unique and traceable are usernames?, in: Privacy Enhancing Technologies, Springer, 2011, pp. 1–17.

[7] D. Irani, S. Webb, K. Li, C. Pu, Large online social footprints–an emerging threat, in: International Conference on Computational Science and Engineering, 2009 (CSE'09), Vol. 3, IEEE, 2009, pp. 271–276.

[8] A. Malhotra, L. Totti, W. Meira Jr., P. Kumaraguru, V. Almeida, Studying user footprints in different online social networks, in: Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), ASONAM '12, IEEE Computer Society, Washington, DC, USA, 2012, pp. 1065–1070.

[9] O. Goga, H. Lei, S. H. K. Parthasarathi, G. Friedland, R. Sommer, R. Teixeira, Exploiting innocuous activity for correlating users across sites, in: Proceedings of the 22Nd International Conference on World Wide Web, WWW '13, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2013, pp. 447–458.

[10] T. Chen, M. A. Kaafar, A. Friedman, R. Boreli, Is more always merrier?: A deep dive into online social footprints, in: Proceedings of the 2012 ACM Workshop on Workshop on Online Social Networks, WOSN '12, ACM, New York, NY, USA, 2012, pp. 67–72.

[11] G. Kontaxis, I. Polakis, S. Ioannidis, E. P. Markatos, Detecting social network profile cloning, in: Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on, IEEE, 2011, pp. 295–300.

[12] K. Krombholz, H. Hobel, M. Huber, E. Weippl, Social engineering attacks on the knowledge worker, in: Proceedings of the 6th International Conference on Security of Information and Networks, ACM, 2013, pp. 28–35.

[13] H. Zhang, D. D. Yao, N. Ramakrishnan, Z. Zhang, Causality reasoning about network events for detecting stealthy malware activities, computers & security 58 (2016) 180–198.

[14] H. Zhang, D. D. Yao, N. Ramakrishnan, Causality-based sense-making of network traffic for android application security, in:

Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security, ACM, 2016, pp. 47–58.

[15] J. R. Finkel, T. Grenager, C. Manning, Incorporating non-local information into information extraction systems by Gibbs sampling, in: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05, Association for Computational Linguistics, Stroudsburg, PA, USA, 2005, pp. 363–370.

[16] M. Kuhn, S. Weston, N. Coulter, M. Culp, C50: C5.0 Decision Trees and Rule-Based Models, R package version 0.1.0-24 (2015).
URL http://CRAN.R-project.org/package=C50

[17] A. Narayanan, V. Shmatikov, De-anonymizing social networks, in: Proceedings of the 30th IEEE Symposium on Security and Privacy (S&P 2009), IEEE Computer Society, 2009, pp. 173–187.

[18] R. Gonzalez, R. Cuevas, R. Motamedi, R. Rejaie, A. Cuevas, Google+ or Google-?: dissecting the evolution of the new OSN in its first year, in: Proceedings of the 22nd international conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2013, pp. 483–494.

[19] M. Edwards, S. Wattam, P. Rayson, A. Rashid, Sampling labelled profile data for identity resolution, in: Proceedings of the IEEE International Conference on Big Data, IEEE, 2016.

[20] M. F. Atig, S. Cassel, L. Kaati, A. Shrestha, Activity profiles in online social media, in: Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, IEEE, 2014, pp. 850–855.

[21] S. Afroz, A. Caliskan-Islam, A. Stolerman, R. Greenstadt, D. McCoy, Doppelgänger finder: Taking stylometry to the underground, in: Proceedings of IEEE Security & Privacy Symposium 2014, 2014.

[22] M. Koppel, J. Schler, Exploiting stylistic idiosyncrasies for authorship attribution, in: Proceedings of IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis, Vol. 69, 2003, p. 72.

[23] Kaspersky Lab, ICS Vulnerabilities Statistics 2015, http://newsroom.kaspersky.eu/fileadmin/user_upload/de/Downloads/PDFs/ICS_Report_Part1_Vulnerabilities.pdf (2015).

[24] B. Green, D. Prince, J. Busby, D. Hutchison, The impact of social engineering on Industrial Control System security, in: Proceedings of the First ACM Workshop on Cyber-Physical Systems-Security and/or PrivaCy, ACM, 2015, pp. 23–29.

[25] D. Gan, L. R. Jenkins, Social networking privacywhos stalking you?, Future Internet 7 (1) (2015) 67–93.

[26] P. Finn, M. Jakobsson, Designing ethical phishing experiments, Technology and Society Magazine, IEEE 26 (1) (2007) 46–58.

[27] K. Lee, J. Caverlee, S. Webb, Uncovering social spammers: Social honeypots + machine learning, in: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '10, ACM, New York, NY, USA, 2010, pp. 435–442.

[28] G. Stringhini, O. Thonnard, That ain't you: Blocking spearphishing through behavioral modelling, in: Detection of Intrusions and Malware, and Vulnerability Assessment - 12th International Conference, DIMVA 2015, Milan, Italy, July 9-10, 2015, Proceedings, 2015, pp. 78–97.

[29] M. C. Kotson, A. Schulz, Characterizing phishing threats with natural language processing, in: Communications and Network Security (CNS), 2015 IEEE Conference on, IEEE, 2015, pp. 308–316.

[30] S. Aggarwal, V. Kumar, S. D. Sudarsan, Identification and detection of phishing emails using natural language processing techniques, in: Proceedings of the 7th International Conference on Security of Information and Networks, Glasgow, Scotland, UK, September 9-11, 2014, 2014.