Social Engineering Detection using Neural Networks

Hanan Sandouka[1], Dr. Andrea Cullen[1] and Ian Mann[2]
[1]*School of Computing, Informatics and Media, Bradford University, UK*
[2]*ECSC Ltd, UK*
h.m.a.sandouka@bradford.ac.uk

***Abstract*- Social Engineering (SE) is considered to be one of the most common problems facing information security today. Detecting social engineering is important because it attempts to secure organisations, consumers and systems from attempts to gain unauthorized access or to reveal some secrets by manipulating employees. The aim of this work is to introduce a new technique for detecting social engineering using neural networks. In this work we have used benchmark data and developed a new technique to extract features that can be used for neural network testing and training. Initial results are encouraging and indicate that machine learning can add an extra layer of security to protect individuals and organisations from social engineering attacks. Future work includes expanding the data set to include additional attack scenarios and benchmark data.**

Keywords-Social Engineering; Neural Network

I. INTRODUCTION

Social engineering (SE) is an emerging science that plays on the trust element of the human nature [1], [2]. It is considered to be an emerging science that attempts to secure organisations, consumers and systems from attempts to gain unauthorized access or to reveal some secrets by manipulating employees. It is becoming increasingly important because of the following reasons:
a) The traditional underestimation of the abilities of social engineers and the general misconception that organisational security systems are very secured by design;
b) The tendency of people to ignore threats thinking that it is a far possibility;
and
c) Human factors, which means that good security policies are mostly not followed because of laziness, fatigue, subjective judgment and the general tendency to be trustful and helpful.

Some of these factors are encouraged by the salesman mentality that "client is always right", which allows social engineers to manipulate the friendly and helpful administrator or helpdesk [3]. Also, social engineers usually take advantage of the victim's emotions as excitement or fear and their good nature that allow them to trust others easily [4]. For example, Dodge, Carver, & Freguson in [5] suggest that many people would give an account username and password in order to get financial interest or valuable prizes.

Social engineering attacks can prove to be very expensive. The significance of these attacks increases with the increase in the number of computer users. It is estimated that 1.2 million computer users in the USA between May 2004 and May 2005 suffered losses caused mainly by phishing. The total cost of these losses reached $929 million. The total losses increased to $3.2 billion between August 2006 and August 2007 and the number of affected users rose to 3.6 million. In the UK as based on recent figures released by the Identity Fraud Steering Committee the Home Office, it is estimated that Identity Theft crimes cost the UK economy around £1.2 billion. Also, losses from phishing almost doubled to 23.2 million pounds in 2005, from 12.2 million pounds in 2004 [3]. In addition, the US department in 2004 concluded that one in three people are more likely to become a victim of SE during their life time [6]. This work aims to use new techniques that could help in minimizing the risk of social engineering attacks by providing the means for an early detection using artificial intelligence (AI) techniques.

This paper is organized as follows: Section 2 provides information about literature review and background of social engineering. Section 3 discusses the existing research and limitations. An explanation for the benchmark data and research methodology is provided in section 4 and 5. Section 6 is devoted to Implementation of the Neural Network for SE Detection and the discussion of the results. Finally, the concluding remarks are provided in Section 7.

978-0-7695-3791-7/09 $26.00 © 2009 IEEE
DOI 10.1109/CW.2009.59

II. LITERATURE REVIEW AND BACKGROUND

Social engineering (SE) may involve psychological and technical aspects in order to gain the victim's trust. It has different definitions, based on the background of the researchers. For some, social engineering is considered to be a human centred field of research. According to Gaudin [7], SE relies on the trusting nature of human beings as it depends on obtaining unauthorized confidential information through impersonating individuals through nontechnical means as he defines SE as "the human side of breaking into a corporate network" [7]. For others it is considered to be a multi-disciplinary research that integrates human and technical–based research. It is defined as "a collection of techniques used to manipulate people into performing actions or divulging confidential information [8].

There are many manifestations for SE:

1. Impersonation/ Pretexting and important user, which is persuading a victim to release information or perform a specific action by creating and using an invented scenario and it is usually done over the telephone. The social engineer usually pretends to be someone who works within the system (an employee). This is a typical example of how social engineers works to obtain the information they need [9].
2. Dumpster Diving: this unique technique depends on throwing away trash contains information that is useless for employees and individuals but important for social engineers. Social engineers usually use this information in footprinting which means that they gather information about their potential target prior to the attack [10].
3. Third Party authorization: the social engineer here tries to convince the victim that he/she has the approval from a trusted third party.
4. Phishing: it is the technique of fraudulently obtaining private information through sending an e-mail that appears to come from a legitimate business such as, a bank, or credit card company, requesting "verification" of information and warning of some consequences if it is not provided. The e-mail usually contains a link to a fraudulent web page that seems to be legitimate and has a form requesting private information.
5. Vishing/IVR or phone phishing: this technique uses "a rogue Interactive voice response (IVR) system to recreate a legitimate sounding copy of a bank or other institution's IVR system. The victim is prompted to call in to the "bank" via a (ideally toll free) number provided in order to "verify" information" [11].
6. Spam mails: e-mails that usually contain attachments with a free download to picture, special program, small gifts...etc. Thapar argues that usually the employee opens these attachments which contain viruses, Trojans and worms that find their way into the systems and networks [12].
7. Instant messages: using chat windows the send online or offline messages that could lead the victim to a phishing site.
8. Popup windows: where a window could appear to the user asking for personal information.
9. Interesting softwares: the social engineer tempts the users to install interesting software on his computer which is usually infected.

and

10. Trojan horse: it is the malware that enables unauthorized access to the user's computers.

III. EXISTING RESEARCH AND LIMITATIONS

One of the main approaches for SE is to assess the vulnerability of security systems in organisations [13]. A penetration testing methodology which consists of a team of experts tasked with exploring and exploiting any discovered vulnerabilities in a targeted information system" is applied. Barrett in [13] suggests using this methodology in order to locate and identify potential victims. This approach is important because it focuses on the human side in SE which is considered to be the weakest link in organisations. Therefore by locating the potential victims through applying the penetration testing methodology there will be a chance to know them and their vulnerabilities and exploiting those vulnerabilities.

Researchers from University of Plymouth karakasiliots, Furnell and Papadaki conducted a study using 179 participants to assess the security threats caused by email-based SE and phishing threats. The participants attended induction sessions to inform them about technical and psychological ploys that are used by SE to deceive end-users. The aim of the study was to determine the degree by which the users can distinguish between legitimate messages and illegitimate ones. It was found that 36% of the participants were able to identify legitimate Emails, while 45% were successful in identifying illegitimate ones. One of the major conclusions for this was that participants could not provide convincing reasons for their decisions [14].

The human factors of SE are discussed in [15]. Organisations should develop their own security culture, which should be embedded in all aspects of activities. A clear strategy is needed to develop this culture. Risk assessment could be implemented for the development of this strategy, which could raise security awareness and result in the development of a human fire wall.

Security awareness represents an important part of SE research. There is increasing evidence that the existing levels of security awareness amongst home users and business is not sufficient to face the increasing threats of SE [16, 14]. Lack of security awareness was a problem during the 1990s as shown in research project [17]. Despite the advances in technical-based techniques to tackle SE in the last decade, security awareness continues to be a major problem. This proves that the lack of awareness. In Dolan [18] and Okenyi and Owens [10] plans for developing security awareness strategies within organisations were introduced. These plans focused on the organisations not the individuals. According to Okenyi and Owens [10] the steps involved in building effective security awareness program are: establish a security policy, identify current training needs, obtain senior management support, determine audiences, define key messages, define available communications vehicles and develop an implementation strategy. The high cost associated with these programs is a major issue. In its first year, these programs can cost 20% of the overall security budget for the organisation. In the following years the cost can drop to 10-15% [19]. In Dolan [18] policies, standards and procedures based on security awareness are proposed to tackle SE. The importance of developing security culture and policies within every organisation is also discussed here. The developed policy should consider the following techniques: passwords policies, vulnerability assessments, data classification, acceptable use policies, background checks, termination processes, incident response, physical security and security awareness training.

An empirical study is carried out by Workman [6] to study the factors behind the success of SE attacks. In this work the relevant literature from management and information security disciplines were surveyed to study their relevance to SE attacks. They reported that there are certain human factors that contribute towards the success of SE attacks such as: threat severity, influence of commitment theory, the likeability to trust and the fear of authority. A questionnaire and observation of behaviours related to the dependant variables were carried out in an organisation that was suffering from SE attacks. They concluded that people with high normative commitment feel obligated to offer information to social engineers in return of free software or gift certificates. Whereas people who are high in continuous commitment tend to provide information for escalating requests. The high affective commitment could also aid social engineers to join a desirable group. Other conclusions were drawn for the threat severity, likability to trust and fear of authority. More information can be found in [6]. The work reported in [14] does not follow a realistic plan. It is true that informing the participants in advance that they will be receiving legitimate and illegitimate messages could increase their awareness. However, the results reported could be true within the context and environment of this experiment but its suitability for real-life work environment needs to be discussed further, especially for longer periods of time. As explained above, the decisions made by the users were subjective and the most of them failed to provide an objective set of rules that they followed to reach their decisions. An AI technique that could be used to represent the experience of users in computerised set of rules could help solve this problem.

Previous research did not largely address the issue of social engineering directly nor did it provide solid theoretical and practical understanding for having successful social engineering intervention. Having solid theoretical understanding of the social engineering phenomena can help all people affected by it to understand this phenomena better and hence plan better intervention techniques that are tailored according to the nature of business.

Existing research seems to lack benchmark data, lack of agreed upon testing techniques, lack of multi-disciplinary research. The lack of benchmark data makes it almost impossible to compare the findings of different researchers and establish the quality of new findings. Existing research is carried out using different data collected from different participants and is carried out within different organisations. In addition, most of existing research focuses on security requirement elicitations, analysis and design issues and neglect testing [20]. The lack of agreed upon testing techniques could be caused by the multi-disciplinary nature of SE, which means that researchers when tackling this topic can be influenced by their background (i.e. computer science, psychology, Human resources, behavioural theorists, etc).

It is worth mentioning that state-of-the-art techniques in computer research such as artificial intelligence and machine learning have not been used yet in SE research [21]. This could be caused by challenges and difficulties associated with the multi-disciplinary nature of applied AI-based research. However, AI techniques offer a novel solution to the SE challenges because of its efficient performance and its automated, real time, and its ability to represent the human experience in computerised learning rules. This can offer the potential to respond to large scale attacks in real time manner and hence increases the organizational security.

IV. BENCHMARK DATA

The data obtained in this work was proposed in [1]. Up to our knowledge this was the first attempt to create a benchmark test set for SE research. The importance of this data comes from the fact that it is the first publicly available benchmark data for social engineering. The data set was artificially generated by Dr. Marcus Rogers, the director of the Cyber Forensics Program in the Dept. of Computer and Information Technology at Purdue University. Dr. Rogers proposed a theoretical solution for the social engineering problem, Social Engineering Defence Architecture (SEDA), which relies on computer systems to analyze phone conversations in real time and determine if the caller is deceiving the receiver [1]. 20 conversation scenarios with nine Social Engineering attacks were chosen for this research, which are located in Appendix A. In these scenarios the employees follow company call policy and request the caller's name, company, and job title in that order. The numbers of the conversations with SE attacks are 3, 4, 7, 9, 10, 11, 15, 18, and 20. In this work we filtered the keywords that are considered to be an indication of social engineering attack. An example is given below:

Scenario # 4:
2222222222 6094660035 9082456896 20050515 1150 #start# This is john candy from apple. I am a computer support tech. to keep your machine up to date I need you to install this on your computer. Thanks bye. #inside# what is your name and who are you with. What is your job title? What can I do for you? Ok I have it installed. You're welcome. Bye.

Scenario	Keyword indicates SE attack	Feature extraction	Attack Result
4	Install	Yes	yes

For more information on the creation and description of this data the reader is referred to [1].

V. RESEARCH METHODOLOGY

In this work, we aim to investigate the feasibility of using AI techniques to provide extra protection from SE attacks in call centres. For this work, we have used the benchmark data of [1] and carried out a number of experiments using neural networks (NN). Figure (1) explains the process, where the two major stages are shown. These stages are the feature extraction stage and the NN learning stage and both are explained in the section to follow.

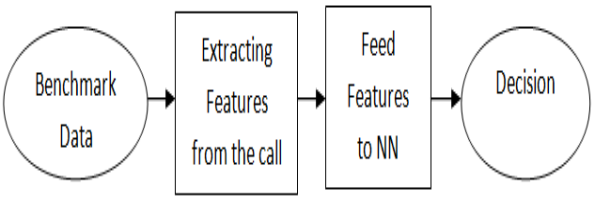

Figure 1. The NN-based system for SE detection

A. Extracting Features From the Phone Call

The aim of this stage is to identify certain attributes/features of the phone call or caller that can help the system identify whether this is an SE attack or not. For this work we have used the data associated with the signature based attacks. The feature extraction process depends on identifying keywords during the call and representing these in numerical training vectors that are arranged to be used for NN learning. An example of the feature extraction process is shown in Table 1.

TABLE 1 THE FEATURE EXTRACTION PROCESS

Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voice identification	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Request sensitive information	0.1	0.1	0.9	0.1	0.1	0.1	0.1	0.1	0.9	0.9	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.9	0.1	0.9
Request password	0.1	0.1	0.1	0.1	0.1	0.1	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.9	0.1	0.1	0.1	0.1	0.1
Installing programs	0.1	0.1	0.1	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
SE attack	0.1	0.1	0.9	0.9	0.1	0.1	0.9	0.1	0.9	0.9	0.9	0.1	0.1	0.1	0.9	0.1	0.1	0.9	0.1	0.9

B. Neural networks (NN)

A neural network (NN) is a computational model that is biologically inspired by the neurones of the human brain [22]. A neural node is the artificial representation of the natural neuron. Each node operates on local information asynchronously without the need for a system clock [23]. Neural networks (NNs) resemble the brain in two respects. Firstly, knowledge is acquired through a learning process. Secondly, neural synaptic weights are used to store the knowledge [24]. The NNs share the following features:

- The outputs of neural nodes are represented by an activation function and it affects the state of all other nodes that it is connected to through weights [25].
- Weights between connected nodes are adjusted during training so as to create a non-linear mapping of the input space to the output space [25].

There are two NNs topologies. The first topology is the feed forward model, where information is fed from the input layer toward the output layer and the connections between nodes do not form cycles. The second topology, which is used in this research, is the feedback model, where cycles exist in feedback NN [22]. The feedback NN is harder to train but has proven to be more efficient for recognition applications such as ours, where we aim to identify SE attacks.

After the feature extraction stage and the creation of the training vectors, several learning experiments using NN were carried out. We have used the MATLAB NN toolbox for this work. A NN with 4 input nodes, 2 hidden nodes and 1 output node was used (see figure 1). 20 scenarios are used in the learning experiments. The aim of the learning experiments was to investigate the feasibility of using NN for the identification of SE attacks. A successful NN learning is indicated by the ability of the NN to converge, which occurs the minimum training error is obtained (see figure 2). For our experiments we have set the training error to be 0.001. The feasibility of using the NN to detect SE can be indicated by the performance of the NN during its training stage. If the NN manages to train then the training error will drop below 0.001. The speed that this takes place also indicates whether NN has managed to detect the patterns behind SE compared to normal calls. For all our experiments, the NN manages to converge fast, usually after some iteration.

VI. PRACTICAL IMPLEMENTATION AND DISCUSSION

The importance of this work comes from the following observations:

1. Up to our knowledge this is the first time the NN has been used to aid in the detection of SE using a benchmark data to test the feasibility of this approach.
2. NN can provide real-time, fast and automatic method to detect SE
3. NN can provide objective performance compared to the subjective performance of humans.ype your main text in 10-point Times, single-spaced.

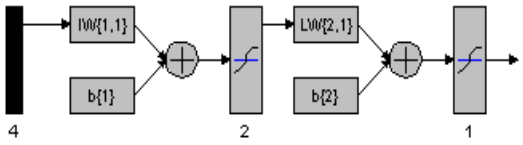

Figure 2. The topology of the NN used

The training trend shows clearly that input features have a clear pattern that can distinguish SE attacks from the normal phone calls. However, our experiments were carried out on limited dataset, which consists of fake scenarios. For the future we need to use real-life data and re-design our NN to handle the new scenarios. Also the NN system should be re-designed to be integrated with an existing call centre and the operators should be encouraged to use it. This technology should be also investigated from the perspective of the Technology Acceptance Model was originally proposed by Davis. According to this model, the two most important determinants of user acceptance of technologies are ease of use and usefulness [26]. These two factors will be investigated further in the future.

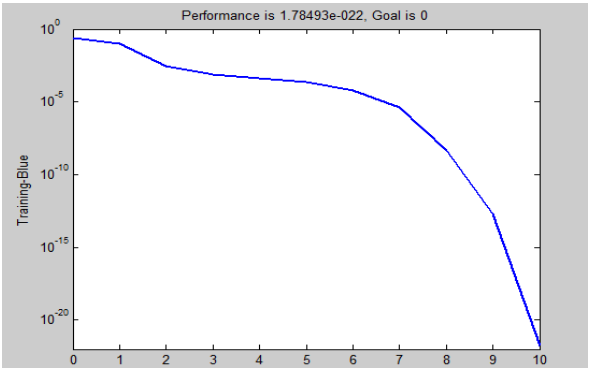

Figure 2. The training error

VII. CONCLUSIONS AND FUTURE WORK

In this paper we investigated the feasibility of using Neural Networks to detect social engineering attacks in Call Centres by identifying certain attributes/features of the phone call or caller that can help the system identify whether this is an SE attack or not. This work is considered to complement previous research in the field of social engineering. The results we got indicate that there is a lot of potential for using this technology in SE. However, a number of modifications and improvements can be made to provide a more efficient framework for detecting SE attacks that does not depend only on filtering the keywords but to extract a set of rules containing the psychological aspects of attackers and/or employees. Also, other elements might be considered such as the time of the call, frequency of calls and length of calls...etc. In addition future research will include investigating this new technology from the TAM perspective and discuss the best methods for implementing this AI-based technology in real-time.

REFERENCES

[1] M.Hoeschele, "Detecting Social Engineering", *MSc Thesis*, Purdue University, 2006.

[2] Mitnick, K. D. and W. L. Simon, *The Art of Deception*, *Controlling the Human Element of Security*, Wiley Publishing, Inc, Indianapolis, Indiana, 2002.

[3] "New Estimate of Cost of Identity Fraud to the UK Economy", http://www.identitytheft.org.uk/cms/assets/cost_of_identity_fraud_to_the_uk_economy_2006-07.pdf. [Last accessed 10/06/2009].

[4] W. Gao and J. Kim, "Robbing the cradle is like taking candy from a baby", *Proceedings of the Annual Conference of the Security Policy Institute (GCSPI),* Amsterdam, The Netherlands, October 2007, pp. 23-37.

[5] R.C. Dodge, C. Carver and A.J. Ferguson, "Phishing for user security awareness", *Computers & Security*, 26, 2007, pp. 73-80.

[6] M. Workman, "Gaining Access with Social Engineering: An Empirical Study of the Threat", *Information Systems Security*, 2007, 16, pp. 315-331.

[7] S. Gaudin, "Social engineering: the human side of hacking", http://itmanagement.earthweb.com/secu/article.php/1040881,2002, [Last accessed 24 July, 2008].

[8] J. Leyden, "Office workers give away passwords for a cheap pen, The Register", http://www.theregister.co.uk/2003/04/18/office_workers_give_away_passwords/?,2003, [Last accessed 24 September 2008].

[9] T.R. Peltier, "Social Engineering: Concepts and Solutions". *EDPACS, informaworld.com*, 33(8), 2006, pp. 1–13.

[10] P. Okenyi and J.O. Owens, "On the Anatomy of Human Hacking", *Information Systems Security*, 16(6), 2007, pp. 302.

[11] A. Smith, "Literature Review", Available on-line: http://www.social-ed.co.uk/SocialEd/lib/attachments/Literature_Review.pdf, 2008, [Last accessed 24 Septemper 2008]

[12] A. Thapar, "Social engineering: An Attack Vector Most Intricate to Tackle". CISSP: Infosec Writers, (106841), 2007.

[13] N. Barret, "Penetration testing and social engineering: hacking the weakest link", *Information Security, Technical Report,* 8 (4), 2003, pp. 56–64.

[14] A. Karakasiliotis, S.M. Furnell and M. Papadaki, "Assessing End-User Awareness of Social Engineering and Phishing", *Proceedings of the 7th Australian Information Warfare and Security Conference, Information Warfare and Security Conference*, 2006, pp. 60-73.

[15] A. Šinigoj, "The Importance of e Security in the Overall e Strategy Of an Organisation", *17th Bled e Commerce Conference e Global Bled*, Slovenia, June 2004, pp. 12–23.

[16] A. Paller, "For Questions: Allan Paller", SANS Institute, Available on-line: http://www.tippingpoint.com/pdf/press/2007/SANSTop20-2007_112707.pdf, 2007, [Accessed: 20/10/08].

[17] T. Greening, "Ask and Ye Shall Receive: A Study in 'Social Engineering'", *ACM Press NY*, 14, 1996, pp. 8-14.

[18] A. Dolan, 2004: "Social Engineering, SANS Reading Room", 2, available online: www.sans.org/rr/catindex.php7cat_id. =51, 2004, [Last accessed 20/10/2008].

[19] P. McBride, "How to Spend a Dollar on Security", *Computerworld*, November 2000.

[20] H. Mouratidis and P. Giorgini, "Security Attack Testing (SAT)—testing the security of information systems at design time", *Information systems*, 32(8), 2007, pp. 1166-1183.

[21] M. Hentea, "Intelligent System for Information Security Management: Architecture and Design Issues", *Issues in Information Science and Information Technology*, 4, 2007, pp. 29-43.

[22] V. Rao and H. Rao, "C++ Neural Networks and Fuzzy Logic", *MIS Press*, New York, 1993.

[23] A. Nirgin, "Neural Networks for Pattern Recognition", *The MIT press*, Cambridge, 1993.

[24] S. Haykin, "Neural Networks: A Comprehensive Foundation", *Macmillan*, New York, 1994.

[25] Y. Pao, "Adaptive Pattern Recognition and Neural Networks", *Addison*, Wesley, 1989.

[26] F. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology", *MIS Quarterly*, 1989.