Towards a development of a Social Engineering eXposure Index (SEXI) using publicly available personal information

W. Shawn Wilkerson
College of Engineering and Computing, Nova Southeastern University, ww364@mynsu.nova.edu

Yair Levy
College of Engineering and Computing, Nova Southeastern University

James Richard Kiper
Federal Bureau of of Investigation, kiper@mynsu.nova.edu

Martha Snyder
College of Engineering and Computing, Nova Southeastern University, smithmt@nova.edu

Abstract
Millions of people willingly expose their lives via Internet technologies every day, and even those who stay off the Internet find themselves exposed through data breaches. Trillions of private information records flow through the Internet. Marketers gather personal preferences to coerce shopping behavior, while providers gather personal information to provide enhanced services. Few users have considered where their information is going or who has access to it. Even fewer are aware of how decisions made in their own lives expose significant pieces of information, which can be used to harm the very organizations they are affiliated with by cyber attackers. While this threat can affect everyone, upper management provides a significantly higher risk due to their level of access to critical data and finances targeted by cybercrime. Thus, the goal of this work-in-progress research is to develop and validate a means to measure exposure to social engineering of 100 executives from Fortune 500 companies. This work-in-progress study will include a mixed methods approach combining an expert panel using the Delphi method, developmental research, and a quantitative data collection. The expert panel will provide a weighted evaluation instrument, subsequently used to develop an algorithm that will form the basis for a Social Engineering eXposure Index (SEXI) using publicly available personal information found on the Internet on these executives, which will help quantify the exposure of each executive. The collected data will be quantitatively evaluated, analyzed, and presented.

Disciplines
Information Security | Management Information Systems | Technology and Innovation

Comments
Official submission to journal and proceedings.

INTRODUCTION

The exposure of personal information on publicly available resources has grown exponentially over the last few years (Acquisti, Brandimarte, & Loewenstein, 2015; Mitnick & Simon, 2002). The Director of the U.S. Central Intelligence Agency (CIA) discovered that no one is safe with the public release of personally identifiable information (PII) of dozens of people from his email account (Federal Bureau of Investigation, 2015; Franceschi-Bicchierai, 2015). Research has shown the primary source of information used in social engineering (SE) attacks originates with the target or their associates (Heartfield & Loukas, 2016; Junger, Montoya, & Overink, 2017; Luo, Brody, Seazzu, & Burd, 2013). Exposure can happen to anybody and enacted by anyone from teenagers to foreign government actors (Federal Bureau of Investigation, 2016; Kopan, 2015). Additionally, studies have indicated that efforts to warn users against sharing their personal information may result in a significant increase of the undesirable behavior even after adjusting privacy settings on social media (Junger et al., 2017; Ku, Chen, & Zhang, 2013; Sutanto, Palme, Tan, & Phang, 2013).

According to Solove (2006), "Exposure involves the exposing to others of certain physical and emotional attributes about a person" (p. 533). Meguerdichian, Koushanfar, Qu, and Potkonjak (2001) define exposure as "a measure of how well an object … can be observed … over a period of time" (p. 139). Publicly available technologies facilitate the exponential growth of PII (Acquisti et al., 2015; Mitnick & Simon, 2002) comprised of social media and other platforms (Krishnamurthy & Wills, 2009; Maynard, Greenwood, Roberts, Windsor, & Bontcheva, 2015). Building upon the discussion of open data by Maynard et al. (2015), publicly available personal information (PAPI) is defined herein as comprising all publicly accessible resources where PII is exposed. The increasingly available PII via PAPI provides social engineers with data, allowing them to be more successful (Acquisti et al., 2015; Mitnick & Simon, 2002), averaging over $100,000 per incident in 2013 (Federal Bureau of Investigation, 2015; Mouton, Leenen, & Venter, 2016).

The research problem that this work-in-progress study will address is the proliferation of SE attacks due to PAPI (Heartfield & Loukas, 2016; Maynard et al., 2015; Mitnick & Simon, 2002). Studies have shown a significant increase in PII exposed via PAPI as well as an overall willingness of Americans to share personal content through self-disclosure, social media, Website personalization, and other public venues (Acquisti et al., 2015; Boyd & Ellison, 2007; Hong & Thong, 2013; Sutanto et al., 2013). Social networking sites, for example, facilitate

this exposure by offering ever-increasing areas of information collection and wider audiences for exposure (Acquisti et al., 2015; Pew Research Center, 2013). Self-disclosure is also on the increase even though 64% of Americans have personally experienced a data breach (Olmstead & Smith, 2017). Additionally, 68% of American adults are using Facebook - an increase of 7% in a single year (Greenwood, Perrin, & Duggan, 2016).

To illustrate the research problem, the SE attack on the CIA Director can serve as a case study encapsulating the effect of PAPI concerning a specific individual through multiple attack vectors. Franceschi-Bicchierai (2015) describes how high school students gathered information concerning the CIA Director from multiple websites: 1) the students obtained the name of the Internet Service Provider 2) from whom they obtained the CIA Director's Social Security number 3) used to request a password reset on a personal email account with another company 4) allowing the students to gain full access to the CIA Director's personal email account, and 5) resulting in the dissemination of the PII of a significant number of CIA agents. Kopan (2015) stated that public figures, such as the CIA Director, are targeted for SE attacks. Mitnick and Simon (2002) made a darker declaration: any inroad into an organization can prove beneficial to the SE agenda.

Most of the literature generalizes PII as any content that can be used to potentially identify an individual (McCallister, Grance, & Scarfone, 2010; Schwartz & Solove, 2011). Schwartz and Solove (2011) argued this stance and propose two additional personal information privacy categories that specify unidentifiable and identifiable personal information. Following the research of Schwartz and Solove (2011), this study defines personally distinguishable information (PDI) as any information that definitively identifies an individual (i.e., images, video, social security number, biometrics) and personally unidentifiable information (PUI) as information that cannot identify a specific individual (i.e., age, date of birth, gender). Henceforth, PII is understood as any information that has the potential of identification, while not falling into the PUI or PDI categories.

The exponential increase of PDI, PII, and PUI available via PAPI is providing the necessary information for successful SE attacks, especially when the target is familiar with the presented information (Acquisti et al., 2015; Heartfield & Loukas, 2016; Neupane, Rahman, Saxena, & Hirshfield, 2015). According to the U.S. Federal Bureau of Investigation (FBI) (2015), Business Email Compromise (BEC) approached $800 million in losses for approximately 7000 U.S. organizations in under two years before August 2015. Additionally, the FBI also logged an increase of 270% of BEC cases over the first eight months of 2015 (Federal Bureau of

Investigation, 2015). Familiarity, distractions, and sleep-deprivation significantly increase the success rate of SE attacks (Heartfield & Loukas, 2016; Neupane et al., 2015). The growth of BEC, SE, and PAPI indicate that current research methodologies may be inadequate (Tetri & Vuorinen, 2013). Thus, additional research is warranted to classify and assess social engineering exposure of individuals, especially for strategic personnel.

Theoretical Background

The privacy chain, defined herein as the flow of PUI/PII/PDI communication between two endpoints, appears to have no lack of supply (Mitnick & Simon, 2002; Tetri & Vuorinen, 2013) or demand (Federal Bureau of Investigation, 2012; Jasper, 2017). Much of personal information originates from people who continue to expose themselves via PAPI even though 64% of America adults have experienced a data breach incident. The ever-expanding flow of PAPI is a constant threat to organizations (Acquisti et al., 2015; Bélanger & Crossler, 2011; Mouton et al., 2016). Junger et al. (2017) indicated the success of an SE attack is due in large part to the availability of personal information. Studies have shown that the number of social media consumers is increasing, with 68% of American adults and 79% of all global Internet users having Facebook accounts (Greenwood et al., 2016). The SE attack of the CIA Director illustrates the significance of PAPI to the successful compromise of an email account, which subsequently led to the exposure of the personal information of many federal agents (Franceschi-Bicchierai, 2015; Kopan, 2015).

Schwartz and Solove (2011) argued for better categorization of personal information to distinguish between PUI, PII, and PDI for legal standing, data storage, long-term consideration, as well as perception. Junger et al. (2017) found that the way people perceive PII in virtual worlds has a direct connection with their mindfulness in the real world. In response to SE threats and the subjectiveness of privacy, many organizations implement security policies and awareness training – though with limited effectiveness (Mitnick & Simon, 2002; Mouton et al., 2016; Schwartz & Solove, 2011; Tetri & Vuorinen, 2013). Current methods are failing to protect organizations from the influx of SE attacks, predict potential targets, provide insight into potential content, or specify possible attack vectors (Heartfield & Loukas, 2016). The relevance of this research is considerable with the documented exponential increase of exposure of PII via PAPI.

The literature has shown that people tend to ignore or not understand privacy policies, appearing unwilling to manage their privacy (Acquisti et al., 2015; Hong

& Thong, 2013). These people comprise those employed by organizations to enforce security policy and provide a cyber defense (Mouton et al., 2016). In a single data breach, one billion user accounts were compromised in late 2016 (Green, 2017). The impact of this is typically underestimated, even though the literature has shown that breached data are often traded in hacker-undergrounds and used in multiple SE attacks (Jasper, 2017). With the release of breached data to a Website, PII is often transformed into PAPI (Franceschi-Bicchierai, 2015; Kopan, 2015). Furthermore, the public release of PII often serves as a flow of continual information used to mount SE attacks through myriad of vectors (Mouton et al., 2016; Tetri & Vuorinen, 2013). Given the documented exponential increase in exposed PII made available to SE via PAPI, the significance of this work-in-progress study is substantial.

Proposed Methodology

The goal of this developmental research is to develop and validate a Social Engineering eXposure Index (SEXI) using publicly available personal information (PAPI) to assist in identifying and classifying SE vulnerabilities, as well as assess SEXI on 100 top US executives. Using a mixed methods research design it is proposed to provide a rating via SEXI indicating the SE exposure a specific individual has. The need for this research is demonstrated by Mitnick and Simon (2002), McCallister et al. (2010), Schwartz and Solove (2011), Sutanto et al. (2013), Tetri and Vuorinen (2013), Heartfield and Loukas (2016), as well as Mouton et al. (2016), who acknowledge the risk associated with SE, the inability of experts to know what information is available to social engineers, the exponential increase of PAPI that can be used to circumvent security methodologies, the low amount of available detail of specific SE events, the growth of SE attack vectors, as well as the lack of a predictive threat system associated with PUI, PII, and PDI.

This work-in-progress study builds upon the work of Tetri and Vuorinen (2013), Acquisti et al. (2015), as well as Heartfield and Loukas (2016). Tetri and Vuorinen (2013) found the availability of PAPI facilitated SE attacks across a variety of vectors. Acquisti et al. (2015) describe the exponential growth of PUI, PII, and PDI across social networks – both in quantity and in the number of willing participants. Additionally, Tetri and Vuorinen (2013) describe the ineffectiveness of current security and research methodologies due to the dearth of specifics for SE attacks combined with the tendency for studies and policies to address only a single lens. Heartfield and Loukas (2016) called for a formal framework that could be used to outline user exposure to SE, while Tetri and Vuorinen (2013) indicated the necessity for future research to investigate the composition and origination of

successful SE attacks. This work-in-progress research would solicit Subject Matter Experts (SMEs) to develop and validate an index of exposure to SE comprised of PUI, PII, PDI available via PAPI for 100 top executives.

The first specific goal of this research study is to gather the requirements for an index of SE exposure using PAPI. The second specific goal of this research study is to develop and assess SMEs approved components and weights for SEXI using the Delphi approach (Ramim & Lichvar, 2014). The third specific goal of this research study is to assess and classify the SEXI of 100 individuals based on the results. The fourth specific goal of this research study is to assess and statistically test for significant mean differences of the SEXI of 100 individuals based on demographical indicators of age, gender, income, marital status, estimated worth, industry, organizational position, philanthropic contributions, and prior military/police experience. The fifth specific goal of this research study is to assess the SEXI of a set of executives, then provide a "best practices" for executives to reduce their threat vector. The research questions that this study will address are:

RQ1: What are the requirements for an index of SE exposure using PAPI?
RQ2: What are the experts' approved components and weights for SEXI using the Delphi expert methodology?
RQ3: How are 100 individuals assessed and classified by SEXI?
RQ4: Are there any statistically significant mean differences of SEXI based on demographic indicators of age, gender, income, marital status, estimated worth, industry, organizational position, philanthropic contributions, and prior military/police experience?
RQ5: Can SEXI assessments provide a set of executives pertinent information by which they can reduce their SE threat vector?

Research indicates that SE attacks are on the increase (Federal Bureau of Investigation, 2015) and that their success (Junger et al., 2017) is often predicated on the availability of PAPI (Acquisti et al., 2015). Embracing Heartfield and Loukas (2016), the development of an exposure index can facilitate the prediction of specific SE targets, as well as the vector composition, typically not provided with current methodologies. Prior research documents the willingness of people to expose their PII (Acquisti et al., 2015; Pew Research Center, 2013, 2015) and the subsequent expense of organizations (Federal Bureau of Investigation, 2015; Mouton et al., 2016).

Conclusions and Discussions

Research indicates that SE attacks are on the increase (Federal Bureau of Investigation, 2015) and that their success (Junger et al., 2017) is often predicated on the availability of PAPI (Acquisti et al., 2015). Embracing Heartfield and Loukas (2016), the development of an exposure index can facilitate the prediction of specific SE targets, as well as the vector composition, typically not provided with current methodologies. Research effectively documents the willingness of people to expose themselves (Acquisti et al., 2015; Pew Research Center, 2013, 2015) and the subsequent expense of organizations (Federal Bureau of Investigation, 2015; Mouton et al., 2016). Thus, the primary aim of this work-in-progress study is the development and validation of a SEXI using PAPI to assist in identifying and classifying SE exposure. People make up the weakest component of any organizational security defense (Mitnick & Simon, 2002), being easily compromised and distractible (Neupane et al., 2015), as well as potentially possessing a representation of privacy associating their virtual life with their physical surroundings (Junger et al., 2017). Organizational efforts to implement security policy and conduct training have shown limited effectiveness (Mouton et al., 2016) and may inadvertently contribute to the problem as research has shown an increase of PII exposure, even after warnings have been issued (Junger et al., 2017; Wolff, 2016). Prior literature has called for a prediction mechanism for potential SE attacks (Heartfield & Loukas, 2016; Mohaisen, Al-Ibrahim, Kamhoua, Kwiat, & Njilla, 2017).

Future Research

This work-in-progress study describes the research plan to develop a set of measures and a single composite index based on criteria identified in the literature. The weights of the criteria and composite index are to be developed using a Delphi approach with SMEs, followed by the development of the SEXI. Data collection and analysis will be performed on 100 executives of US organizations, where anonymized data will be analyzed and reported.

Future studies are warranted to increase the validity of the SEXI benchmarking index. Also, more research is required to expand the sample size, and the use of other populations to enhance the generalizability of the measure developed. While this work-in-progress study is concentrating on 100 executives of US organizations, future research could include other groups of individuals who may be exposed to significant SE attacks, such as law enforcement officers, government employees, civilian contractors to the government, as well as others who may be a point of

contact for proprietary information at various organizations. Moreover, future work can assess privacy practices related to social media and personalization, or even a longitudinal study on the impact of awareness of these individuals to their SEXI benchmarking index value over time and what actions they took to mitigate their exposure to SE attacks. Another area of future research can include the selection of a population based on demographic indicators (i.e., gender, age, education) to determine if there are any significance difference levels of the SEXI benchmarking index based on these indicators. Due to the increasing number of SE attacks, future research could include assessing the composite SEXI of an organization or sections thereof. Additionally, research into the values of the SEXI benchmarking index when compared based on culture may provide further insight.

Limitations

This work-in-progress study has several limitations. In its current state, the SEXI benchmarking index is based on the foundational literature, and the feedback, validation, and adjustments are needed from the Delphi approach with SMEs who may provide some important adjustments that were not previously reported in literature. The second limitation is the set of measures combined to form SEXI. Given that cyber attacks and SE attacks in particular are changing over time, the SEXI benchmarking index is based on the current SE threat vector, techniques, or approaches, while we envision it requiring more adjustments in the future when new approaches to SE, social media security and privacy settings, as well as identity theft emerge. The third limitation is the reliance on an American group of experts for the SME panel to establish the instrument. While it is the limitation of the current work-in-progress study, we believe that a broader population of SMEs with also more international participation that is well represented around the globe may provide more generalizability to the relative weights, criteria, and measures.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509-514. doi:10.1126/science.aaa1465

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly, 35*(4), 1017-A1036.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230. doi:10.1111/j.1083-6101.2007.00393.x

Federal Bureau of Investigation. (2012). Internet social networking risks. *Counterintelligence.* Retrieved from https://www.fbi.gov/file-repository/internet-social-networking-risks-1.pdf/view

Federal Bureau of Investigation. (2015). Business e-mail compromise. *Stories.* Retrieved from https://www.fbi.gov/news/stories/2015/august/business-e-mail-compromise/business-e-mail-compromise

Federal Bureau of Investigation. (2016). Iranians charged with hacking U.S. financial sector. Retrieved from https://www.fbi.gov/news/stories/iranians-charged-with-hacking-us-financial-sector/iranians-charged-with-hacking-us-financial-sector

Franceschi-Bicchierai, L. (2015). Teen hackers: A '5-year-old' could have hacked into CIA Director's emails. Retrieved from https://motherboard.vice.com/read/teen-hackers-a-5-year-old-could-have-hacked-into-cia-directors-emails

Green, N. (2017). Standing in the future: The case for a substantial risk theory of "injury in fact" in consumer data breach class actions. *Boston College Law Review, 58*(1), 287-351.

Greenwood, S., Perrin, A., & Duggan, M. (2016, November 11, 2016). Social media update 2016. Retrieved from assets.pewresearch.org/wp-content/uploads/sites/14/2016/11/10132827/PI_2016.11.11_Social-Media-Update_FINAL.pdf

Heartfield, R., & Loukas, G. (2016). A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks. *ACM Computing Surveys (CSUR), 48*(3), 37. doi:10.1145/2835375

Hong, W., & Thong, J. Y. L. (2013). Internet privacy concerns: An integrated conceptualization and four empirical studies. *MIS Quarterly, 37*(1), 275-298.

Jasper, S. E. (2017). U.S. cyber threat intelligence sharing frameworks. *International Journal of Intelligence and CounterIntelligence, 30*(1), 53-65. doi:10.1080/08850607.2016.1230701

Junger, M., Montoya, L., & Overink, F. J. (2017). Priming and warnings are not effective to prevent social engineering attacks. *Computers in Human Behavior, 66*, 75-87. doi:10.1016/j.chb.2016.09.012

Kopan, T. (2015). CIA Director John Brennan 'outraged' by hack of his emails. Retrieved from http://www.cnn.com/2015/10/27/politics/john-brennan-email-hack-outrage/

Krishnamurthy, B., & Wills, C. E. (2009). *On the leakage of personally identifiable information via online social networks.* Proceedings of the 2nd ACM workshop on Online social networks, Barcelona, Spain.

Ku, Y.-C., Chen, R., & Zhang, H. (2013). Why do users continue using social networking sites? An exploratory study of members in the United States and Taiwan. *Information & Management, 50*(7), 571-581. doi:10.1016/j.im.2013.07.011

Luo, X. R., Brody, R., Seazzu, A., & Burd, S. (2013). Social engineering: the neglected human factor for information security management. *Managing Information Resources and Technology: Emerging Applications and Theories: Emerging Applications and Theories*. doi:10.4018/irmj.2011070101

Maynard, D., Greenwood, M. A., Roberts, I., Windsor, G., & Bontcheva, K. (2015). Real-time social media analytics through semantic annotation and linked open data. *Proceedings of the ACM Web Science Conference*, 1-2. doi:10.1145/2786451.2786500

McCallister, E., Grance, T., & Scarfone, K. (2010). *Guide to protecting the confidentiality of personally identifiable information (PII)*. (SP 800-122). Washington, DC: National Institute of Standards and Technology (NIST) Retrieved from http://ws680.nist.gov/publication/get_pdf.cfm?pub_id=904990.

Meguerdichian, S., Koushanfar, F., Qu, G., & Potkonjak, M. (2001). Exposure in wireless ad-hoc sensor networks. *Proceedings of the 7th annual international conference on Mobile computing and networking*, 139-150. doi:10.1145/381677.381691

Mitnick, K. D., & Simon, W. L. (2002). *The art of deception: Controlling the human element of security*. New York, NY: John Wiley & Sons.

Mohaisen, A., Al-Ibrahim, O., Kamhoua, C., Kwiat, K., & Njilla, L. (2017). Rethinking information sharing for actionable threat intelligence. *arXiv preprint arXiv:1702.00548*.

Mouton, F., Leenen, L., & Venter, H. S. (2016). Social engineering attack examples, templates and scenarios. *Computers & Security, 59*, 186-209. doi:10.1016/j.cose.2016.03.004

Neupane, A., Rahman, M. L., Saxena, N., & Hirshfield, L. (2015). A multi-modal neuro-physiological study of phishing detection and malware warnings. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 479-491. doi:10.1145/2810103.2813660

Olmstead, K., & Smith, A. (2017, January 26, 2017). Americans and cybersecurity. *Pew Research Center*. Retrieved from assets.pewresearch.org/wp-content/uploads/sites/14/2017/01/26102016/Americans-and-Cyber-Security-final.pdf

Pew Research Center. (2013). Social networking fact sheet. Retrieved from pewinternet.org/fact-sheets/social-networking-fact-sheet/

Pew Research Center. (2015). U.S. smartphone use in 2015.

Ramim, M. M., & Lichvar, B. T. (2014). Eliciting expert panel perspective on effective collaboration in system development projects. *Online Journal of Applied Knowledge Management, 2*(1), 122-136.

Schwartz, P. M., & Solove, D. J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review, 86*(6), 1814.

Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review, 154*, 477+.

Sutanto, J., Palme, E., Tan, C.-H., & Phang, C. W. (2013). Addressing the personalization-privacy paradox: An empirical assessment from a field experiment on smartphone users. *MIS Quarterly, 37*(4), 1141-A1145.

Tetri, P., & Vuorinen, J. (2013). Dissecting social engineering. *Behaviour & Information Technology, 32*(10), 1014-1023. doi:10.1080/0144929X.2013.763860

Wolff, J. (2016). Perverse effects in defense of computer systems: When more Is less. *Journal of Management Information Systems, 33*(2), 597-620. doi:10.1080/07421222.2016.1205934