UNIVERSITI PUTRA MALAYSIA

SmiDCA: Smishing attack detection for mobile computing on smishing dataset

DAHAH AHMED HAIDARAH HASAN

FSKTM 2019 15

SmiDCA: Smishing Attack Detection For Mobile Computing on Smishing Dataset

By

DAHAH AHMED HAIDARAH HASAN

Dissertation Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Information Security

Jan 2019

DEDICATIONS

To the amiability of difficult days and the moon of dark nights . . . my precious parent.

To each of who has taught me or gave me advice . . . teachers.

Dedicate this work . . .

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Information Security

SmiDCA: Smishing Attack Detection For Mobile Computing on Smishing Dataset

By

DAHAH AHMED HAIDARAH HASAN

JANUARY 2019

Supervisor: Zurina Mohd Hanapi, Assoc. Prof. Dr.

Faculty: Computer Science and Information Technology

Nowadays nearly everyone is using mobile computer/devices such as smart-phones and laptops to conduct their business transactions and for social purposes. While this trend has significantly transformed working and personal lifestyles worldwide, it has also led to serious concerns about threats to security and privacy among individuals as well as organizations. One of the most widespread security threats is phishing attacks launched for the purpose of stealing certain sensitive information of victims and then abusing this information to illegally obtain confidential data. There are many types of phishing attack such as social phishing, spear-phishing, pharming, and smishing. Recently Joo et al. (2017) proposed an improved security prototype to detecting Smishing attack on mobile computing known as S-Detector. Their model is able to distinguish between normal SMS message and phishing. However Goel and Jain (2017a) claimed that S-Detector does not address three SMS security message features. First, S-Detector cannot not check for login page within the SMS message. Second, it is not efficient in detecting self-answering messages and Lastly,

text normalization is not achieved. To solve these issues (Sonowal and Kuppusamy, 2018) propose new technique called SmiDCA. In this research, we re-implement SmiDCA using dataset called smishing dataset for Harm ans Spam (Almeida, 2017). The re-implement SmiDCA technique is analyzed SMS messages and extracted the security features of SMS to detect the smishing SMS messages efficiently.

ACKNOWLEDGMENTS

First and foremost, praise be to Allah through whose mercy (and favors) all good things are accomplished. ("My Lord, increase me in knowledge." . Surat Taha 20:114).

I would like to thank my parents. I am deeply indebted to them for their unconditional support and sacrifice for so many years. I would like to express the deepest appreciate and sincerest gratitude to my supervisor, Assoc. Prof. Dr. Zurina Mohd Hanapi, for her patience and truthful guidance through all the steps of research and writing dissertation. I attribute the level of my Master degree to her encouragement and effort she spent for helping me accomplishing this work.

I thank and appreciate the valuable effort of my lecturers at Faculty of Computer Science and Information Technology (FSKTM), Universiti Putra Malaysia (UPM), Malaysia.

Finally, I owe many thanks to my wife for her love, dedication, help, and encouragement in those critical moments along this journey. Words are not enough to express my gratitude.

This dissertation was submitted to the Information Security Department, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of **Master of Information Security**.

Zurina Mohd Hanapi, PhD
Assoc. Prof. Dr.
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia

Date: Jan 1, 2019

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

(Signature)

DAHAH AHMED HAIDARAH HASAN
(GS50436)
Date: Jan 1, 2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

This chapter commences with an overview of the research topic and explains the motivation for this work. The problem statement is then identified followed by the, research objectives, and scope. At the end of this chapter the organization of the dissertation is provided.

1.1 Background and Motivation

Today, mobile devices particularly smartphones are being increasingly used or a wide range of functionalities for work and social purposes. Meanwhile, it has been noted that this mobile phishing URL click rate has increased by 85% year-on-year as reported by Patrick (2018). The threat is not confined to email, SMS phishing attacks have also been regularly reported with over 25% occurrence. As reported by Patrick (2018), targets click malicious links from spoofed phone numbers that falsely appear to be from the victim's area code. Attackers now take advantage of SMS and MMS as a means of phishing, while some of today's popular social media applications and messaging platforms, including WhatsApp, Facebook Messenger, and Instagram are not spared. However, the detection of mobile phishing attack is a problem different from desktop phishing due to the different designs of both. Moreover, the capability to identify mobile phishing attack with high accuracy is a crucial research need and a challenging one considering that not much work has been done in this field. Many anti-phishing solutions for mobile devices have been proposed to date, but a fully effective and reliable solution is still to be found.

The aim of this dissertation is therefore to highlight phishing attacks on mobile systems and to propose a technique to detect the attacks through SMS sent to mobile phones.

1.2 Problem Statement

According to a review conducted by (Jain and Gupta, 2018) the S-Detector (Joo et al., 2017) does not meet three of the security features of SMS message. First S-Detector cannot check for login page within the SMS message. Second, this model is not efficient in detecting self-answering messages. Lastly, text normalization is not achieved. It is important to check if the SMS message contains a login page within. This is to maintain the confidentiality of user's information in case the message is a phishing attack. This model is not efficient in detecting self-answering messages, and this will allow attacker to take over the systems. Eventually, therefore, the text normalization feature of SMS is very crucial for detecting the phishing of SMS message, and ignoring this feature reduces the accuracy of smishing detection technique (Goel and Jain, 2017a). One solution is introduced by Sonowal and Kuppusamy (2018) called SmiDCA to solve these three security features however their solution has difficult in implementing on different platform like Windows and needs to be analyzed.

1.3 Research Objectives

The objective of this research is to re-implement anti-Smishing(SmiDCA) in order to improve the accuracy of Smishing attack using smishing dataset for Ham and Spam (Almeida, 2017).

1.4 Research Scope

Among the most important methods for communicate in mobile environment are SMS (short message services), and MMS (multimedia message services which include pictures and videos). However, in this study, the main focus is SMS. An SMS comprises up to 160 characters in a single message, and contains SMS features. Fur-

thermore, testing is based on simulation only. Smishing can merge phishing attacks, mobile environment, and techniques they used. Figure 1.1 illustrates the relationship between phishing detecting techniques, phishing attacks and mobile environment.

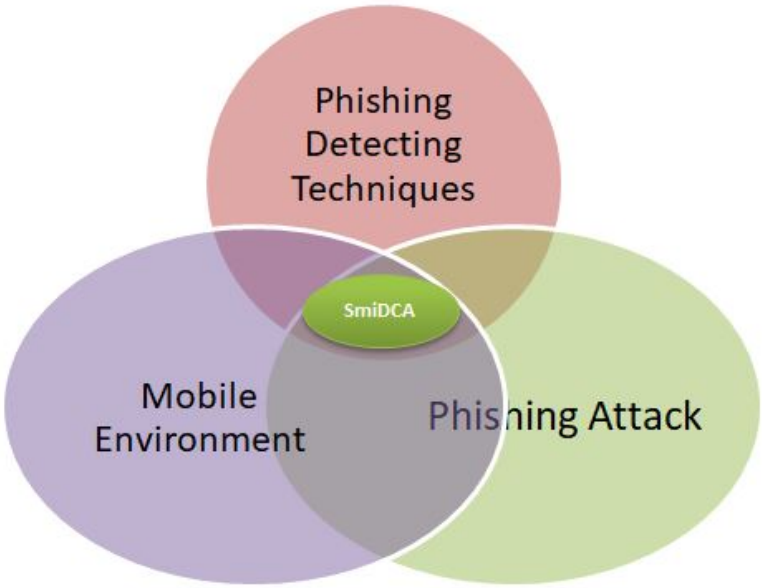

Figure 1.1
Research Scope

The scope of study is the merge between all phishing detection techniques, phishing attacks and mobile environment, in the process of developing the new SmiDCA technique.

1.5 Organization of Dissertation

This research comprises of five chapters. Chapter 1 is the introduction chapter covers the background of the study, identifies and discusses the problem statement, research objectives, explains the scope of research, ends with a brief description of the organization of the dissertation. Chapter 2 introduces the literature review on mobile computing concepts, and presents three main areas that focus on mobile computing(architecture, and application), issues on mobile computing, and mobile-base social engineering. Each of these concepts is comprehensively highlighted and journals with

studies related to the research topic are identified. In reviewing the literature on in mobile-base social engineering, gaps and challenges related to problem statement are identified and discussed. Chapter 3 focuses on the research methodology and presents the SmiDCA framework in detail. What follows are the proposed research technique, and implementation by employing python version anaconda 3.5 through the Kali Linux environment. The results are presented and discussed in Chapter 4 while Chapter 5 offers the conclusions of the study and makes recommendations for further related research.

REFERENCES

Ahmed, I., Ali, R., Guan, D., Lee, Y.-K., Lee, S. and Chung, T. 2015. Semi-supervised learning using frequent itemset and ensemble learning for SMS classification. *Expert Systems with Applications* 42 (3): 1065–1073.

Akinyelu, A. A. and Adewumi, A. O. 2014. Classification of phishing email using random forest machine learning technique. *Journal of Applied Mathematics* 2014.

Aleroud, A. and Zhou, L. 2017. Phishing environments, techniques, and countermeasures: A survey. *Computers & Security* 68: 160–196.

Almeida, T. 2017. Ham and spam dataset .

Almeida, T., Hidalgo, J. M. G. and Silva, T. P. 2013. Towards sms spam filtering: Results under a new dataset. *International Journal of Information Security Science* 2 (1): 1–18.

Almeida, T. A., Hidalgo, J. M. G. and Yamakami, A. 2011. Contributions to the study of SMS spam filtering: new collection and results. In *Proceedings of the 11th ACM symposium on Document engineering*, 259–262. ACM.

Almeida, T. A., Silva, T. P., Santos, I. and Hidalgo, J. M. G. 2016. Text normalization and semantic indexing to enhance instant messaging and SMS spam filtering. *Knowledge-Based Systems* 108: 25–32.

Alsaleh, M., Alarifi, A., Al-Quayed, F. and Al-Salman, A. 2015. Combating comment spam with machine learning approaches. In *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on*, 295–300. IEEE.

Alzahrani, A. J. and Ghorbani, A. A. 2014. SMS mobile botnet detection using a multi-agent system: research in progress. In *Proceedings of the 1st International Workshop on Agents and CyberSecurity*, 2. ACM.

Bijalwan, V., Kumar, V., Kumari, P. and Pascual, J. 2014. KNN based machine learning approach for text and document mining. *International Journal of Database Theory and Application* 7 (1): 61–70.

Bonneau, J., Herley, C., Van Oorschot, P. C. and Stajano, F. 2015. Passwords and the evolution of imperfect authentication. *Communications of the ACM* 58 (7): 78–87.

Bottazzi, G., Casalicchio, E., Cingolani, D., Marturana, F. and Piu, M. 2015. MP-Shield: a framework for phishing detection in mobile devices. In *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*, 1977–1983. IEEE.

Chaudhry, J. A., Chaudhry, S. A. and Rittenhouse, R. G. 2016. Phishing attacks and defenses. *International Journal of Security and Its Applications* 10 (1): 247–256.

Chekina, L., Shapira, B., Mimran, D., Elovici, Y. and Peylo, C. 2016, System for detection of mobile applications network behavior-netwise, uS Patent 9,369,476.

Choudhary, N. and Jain, A. K. 2017, In Advanced Informatics for Computing Research, In *Advanced Informatics for Computing Research*, 18–30, Springer, 18–30.

Curnyn, J. 2014, Anti-phishing system, uS Patent 8,635,666.

Eberhardt III, J. S., Radano, T. A. and Peterson, B. E. 2016, Application of machine learned Bayesian networks to detection of anomalies in complex systems, uS Patent 9,349,103.

El-Alfy, E.-S. M. and AlHasan, A. A. 2016. Spam filtering framework for multimodal mobile communication based on dendritic cell algorithm. *Future Generation Computer Systems* 64: 98–107.

Galán-García, P., Puerta, J. G. d. l., Gómez, C. L., Santos, I. and Bringas, P. G. 2016. Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying. *Logic Journal of the IGPL* 24 (1): 42–53.

Goel, D. and Jain, A. K. 2017a. Mobile phishing attacks and defence mechanisms: State of art and open research challenges. *Computers & Security* .

Goel, D. and Jain, A. K. 2017b. Smishing-Classifier: A Novel Framework for Detection of Smishing Attack in Mobile Environment. In *International Conference on Next Generation Computing Technologies*, 502–512. Springer.

Han, B., Cook, P. and Baldwin, T. 2013. Lexical normalization for social media text. *ACM Transactions on Intelligent Systems and Technology (TIST)* 4 (1): 5.

He, D., Chan, S. and Guizani, M. 2015. Mobile application security: malware threats and defenses. *IEEE Wireless Communications* 22 (1): 138–144.

Hidalgo, J. M. G., de Buenaga Rodríguez, M. and Pérez, J. C. C. 2005. The role of word sense disambiguation in automated text categorization. In *International Conference on Application of Natural Language to Information Systems*, 298–309. Springer.

Ismael, O. A. 2016, Optimized resource allocation for virtual machines within a malware content detection system, uS Patent 9,495,180.

Jain, A. K. and Gupta, B. 2018. Rule-Based Framework for Detection of Smishing Messages in Mobile Environment. *Procedia Computer Science* 125: 617–623.

Jiang, L., Wang, S., Li, C. and Zhang, L. 2016. Structure extended multinomial naive Bayes. *Information Sciences* 329: 346–356.

Joe, I. and Shim, H. 2010. An SMS spam filtering system using support vector machine. In *International Conference on Future Generation Information Technology*, 577–584. Springer.

Joo, J. W., Moon, S. Y., Singh, S. and Park, J. H. 2017. S-Detector: an enhanced security model for detecting Smishing attack for mobile computing. *Telecommunication Systems* 66 (1): 29–38.

Karami, A. and Zhou, L. 2014. Improving static SMS spam detection by using new content-based features .

Liu, Z., Lin, W., Li, N. and Lee, D. 2005. Detecting and filtering instant messaging spam-a global and personalized approach. In *Secure Network Protocols, 2005.(NPSec). 1st IEEE ICNP Workshop on*, 19–24. IEEE.

Longe, O. B., Abdulganiyu, A., Longe, F. and Adegoke, K. 2012. A Prototype Scalable System for Secured Bulk SMS Delivery on Mobile Networks. *International Journal of Advanced Research in Computer Science* 3 (1).

Maroof, U. 2010. Analysis and detection of SPIM using message statistics. In *Emerging Technologies (ICET), 2010 6th International Conference on*, 246–249. IEEE.

Mathew, K. and Issac, B. 2011. Intelligent spam classification for mobile text message. In *Computer Science and Network Technology (ICCSNT), 2011 International Conference on*, 101–105. IEEE.

Meier, J. 2008. Mobile Application Architecture Guide .

Modupe, A., Olugbara, O. O. and Ojo, S. O. 2014, In Transactions on Engineering Technologies, In *Transactions on Engineering Technologies*, 671–686, Springer, 671–686.

Narudin, F. A., Feizollah, A., Anuar, N. B. and Gani, A. 2016. Evaluation of machine learning classifiers for mobile malware detection. *Soft Computing* 20 (1): 343–357.

Nuruzzaman, M. T., Lee, C. and Choi, D. 2011. Independent and personal SMS spam filtering. In *Computer and Information Technology (CIT), 2011 IEEE 11th International Conference on*, 429–435. IEEE.

Patrick. 2018. Phishing attacks against mobile .

Rossi, R. G., de Andrade Lopes, A. and Rezende, S. O. 2016. Optimization and label propagation in bipartite heterogeneous networks to improve transductive classification of texts. *Information Processing & Management* 52 (2): 217–257.

Sahu, S. and Mehtre, B. M. 2015. Network intrusion detection system using J48 Decision Tree. In *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on*, 2023–2026. IEEE.

Silva, R. M., Almeida, T. A. and Yamakami, A. 2017. MDLText: An efficient and lightweight text classifier. *Knowledge-Based Systems* 118: 152–164.

Skudlark, A. 2014. Characterizing SMS spam in a large cellular network via mining victim spam reports .

Sonowal, G. and Kuppusamy, K. 2018. SmiDCA: An Anti-Smishing Model with Machine Learning Approach. *The Computer Journal* .

Titonis, T. H., Manohar-Alers, N. R. and Wysopal, C. J. 2017, Automated behavioral and static analysis using an instrumented sandbox and machine learning classification for mobile security, uS Patent 9,672,355.

Uysal, A. K., Gunal, S., Ergin, S. and Gunal, E. S. 2012. A novel framework for SMS spam filtering. In *Innovations in Intelligent Systems and Applications (INISTA), 2012 International Symposium on*, 1–4. IEEE.

Wu, L., Du, X. and Wu, J. 2014a. MobiFish: A lightweight anti-phishing scheme for mobile phones. In *Computer Communication and Networks (ICCCN), 2014 23rd International Conference on*, 1–8. IEEE.

Wu, L., Du, X. and Wu, J. 2016. Effective defense schemes for phishing attacks on mobile computing platforms. *IEEE Transactions on Vehicular Technology* 65 (8): 6678–6691.

Wu, Q., Ye, Y., Zhang, H., Ng, M. K. and Ho, S.-S. 2014b. ForesTexter: an efficient random forest algorithm for imbalanced text categorization. *Knowledge-Based Systems* 67: 105–116.

Yadav, K., Kumaraguru, P., Goyal, A., Gupta, A. and Naik, V. 2011. SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering. In *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, 1–6. ACM.

Zhang, J. and Wang, Y. 2012. A real-time automatic detection of phishing URLs. In *Computer Science and Network Technology (ICCSNT), 2012 2nd International Conference on*, 1212–1216. IEEE.

Zhang, Y., Hong, J. I. and Cranor, L. F. 2007. Cantina: a content-based approach to detecting phishing web sites. In *Proceedings of the 16th international conference on World Wide Web*, 639–648. ACM.