A review for a system to detect and notify phishing attacks in mobile phones.

DOI: 10.29322/IJSRP.12.08.2022.p12830
http://dx.doi.org/10.29322/IJSRP.12.08.2022.p12830

Paper Received Date: 19th July 2022
Paper Acceptance Date: 06th August 2022
Paper Publication Date: 16th August 2022

KKH Nethmini
Department of Computer Science
General Sir John Kotelawala Defence University
Ratmalana, Sri Laanka.
36-cs-0020@kdu.ac.lk

NT Jayatilake
Department of Computer Engineering
General Sir John Kotelawala Defence University
Ratmalana, Sri Laanka.
jayatilakan@kdu.ac.lk

Thushara Weerawardane
Department of Computer Engineering
General Sir John Kotelawala Defence University
Ratmalana, Sri Laanka.
tlw@kdu.ac.lk

***Abstract*— Because of the advancement of technology, attackers have shifted their focus away from personal computers and onto smartphones, making mobile security a big concern these days. Furthermore, as technology advances, people are becoming more interested in cell phones. Smishing is a cyber security attack that uses the short message service to steal mobile users' personal information. Attackers have taken advantage of users' trust in their smart gadgets to carry out various mobile security exploits such as smishing. If there is a system or approach to identify these kinds of malicious attacks, it is very useful. This paper includes a survey conducted to get a clear idea on awareness of smishing attacks of people in society and identify the need for a system to detect smishing attacks. According to the survey, although a person with good experience with smishing attacks can detect a smishing attack by looking at the features, people who do not have proper knowledge and experience about smishing attacks may not recognize Smishing attacks properly. Moreover, the paper includes a literature survey along with summarized existing systems to detect smishing attacks. These systems have used algorithms and approaches. Some of them are machine learning algorithms, Random Forest algorithm, Feature-based technique, Optical Character Recognition, Tag & APK check, Rule-based approach, Naïve Bayes classifier, Heuristic approach, Support Vector Machine, Rank correlation algorithms, Decision tree, and Ada boost classifier. If a system or approach can work with high accuracy and efficiency that approach or system can be identified as a successful one.**

Keywords*— Smishing, Algorithms, Optical Character Recognition (OCR), Machine learning, Mobile phones, Mobile security

I. INTRODUCTION

Phishing is an online scam in which fraudulent e-mails, text messages, ads, and other methods are used to steal sensitive information. Phishing is a technology used by criminals to obtain sensitive information such as usernames or passwords. It's a social engineering technique. Malicious software such as ransomware, spyware, or a virus will be placed on your device if you open an attachment or click on a phishing link in these emails or messages that look to be from someone you know and trust. It normally takes place behind the scenes, so the average individual is unaware. Phishing was the most common sort of cybercrime in 2020, according to the FBI, and phishing instances nearly doubled in frequency, from 114,702 in 2019 to 241,324 in 2020 [1]. The main types of phishing attacks are email phishing, spear phishing, Smishing, Vishing, HTTP phishing, etc.

Mobile phishing is a new danger that targets financial institutions, online retailers, and social networking sites mobile users. As Information Technology develops and grows, mobile phones have become an irreducible part of ourdaily lives. These devices are becoming more popular among consumers because of their extended battery life, tiny size, and portability. With the increased use of mobile cellphones, the security risk associated with these devices has also increased.Due to the small display of mobile phones, lack of understanding among mobile users, less secure user behaviors, and habits of users to enter their credentials frequently, mobile smartphones appear to be an easier targetfor attackers when compared to desktop devices.

Smishing is the type of phishing that can affect mobile phones. Smishing word comes by combining SMS and phishing words. Smishing is a popular phishing attack that sends text messages as it launches an attack on a cybersecurity unit in the company. Smishing is unknown to less than 35% of the population. It might be difficult for mobile users, particularly elderly users, to determine the legitimacy of a text message. One of the reasons SMS-based fraud is so successful is because of this [2]. The attacker sends a text message to the victim's phone, convincing them to click a link in the message. When the user clicks this link, one of two things happens: the user is taken to a phishing page where they are deceived into entering their login credentials, or the device is silently infected with surveillance malware. The attacker's ultimate goal is to get unauthorized access to the device's personal, sensitive, and corporate data stored and accessed. Because attackers may target a large number of users with a cheaper SMS subscription, text messages are chosen by attackers to target victims. In the case of text messages, smishers or attackers get a higher response rate from consumers than in the case of e-mails.

By 2021, smishing attacks have increased at a very high rate. Although the term "smishing" was coined in 2006, this type of assault was mostly unknown until recently. SMS- based scams increased by 328 percent in the middle of 2020, according to Proofpoint [3]. When the COVID-19 pandemic broke out, officials started sending out SMS messages regarding lockdowns, contact tracking, and vaccine alternatives. A flood of bogus SMS messages arose as a result of this. According to NextCaller, during the first two weeks of the nationwide quarantine period, 44 percent of US Americans reported a spike in scam phone calls and text messages [4].

This literature survey provides the methods which have been used by other researchers to detect these kinds of smishing attacks. Machine learning classification methods, Heuristic-based methods are used to extract the most efficient

features of smishing attacks using keywords inside a message. Algorithms like Backpropagation can be used to classify the messages. To analyze the malicious behavior of these kinds of attacks we can use tag check and APK download check. To find out whether an URL is phishing or not can use an algorithm based on heuristic anti-phishing technical approaches.

The remaining sections of this paper are as follows. The literature review includes what researchers have done so far to identify smishing attacks and what algorithms have been used for each system. The discussion section includes a survey done by using people in society to get an idea about their smishing awareness and collect their smishing attack experiences. Moreover, the discussion includes an experimental evaluation of algorithms and features used in existing systems. Finally, the paper includes a conclusion and further work.

II. Literature Review

This chapter includes a summary of some relevant research work done on detecting mobile phishing attacks, within the past 2 decades.

D. Soni and S. Mishra in 2019 have done research on a Content-Based Approach for detecting Smishing in Mobile Environment [5]. The authors proposed a system that has the ability to categorize text messages by using message contents and the behavior of the URL. They classified the message using a machine learning system based on malicious phrases in the message. They also implemented techniques like form tag check and APK download check to investigate the URL's harmful activity. Based on the results of the detection procedures, text messages will be categorized as malicious or non-malicious. Python is the language that is used to implement this system. Moreover, this system has the ability to identify a smishing attack when they receive to mobile phone and the message can be discarded if it is a smishing.

According to the research on Rule-based framework for Detection of Smishing Messages in Mobile Environment by A. K. Jain and B. B. Gupta in 2017 [6], it indicates that how the smishing attack impact the Short Message Service to steal personal details, bank credentials, login credentials, etc. As a solution authors have implemented a rule-based approach to identify smishing attacks. They used text normalization techniques to this system to convert SMS text messages to a standard form to get a better solution. After evaluating the performance of this system, they gained a 99% true negative rate for this system. Moreover, this system has the ability to detect zero-day attacks also. According to this evaluation, this system has a very efficient detection.

In 2014, L. Wu, X. Du, and J. Wu proposed a system named MobiFish [7] which is a lightweight anti-phishing scheme for mobile phones. By comparing the actual identification to the identity claimed by the web pages and Apps, MobiFish confirms the validity of web pages and apps. MobiFish solves this problem by employing optical character recognition (OCR), which can accurately extract text from a screenshot of a mobile login window, allowing the claimed identity to be validated. As the authors say MobiFish has been implemented in a Google Nexus 4 smartphone which has the Android 4.2 OS. Moreover, MobiFish can effectively identify phishing attacks and defend against them.

J. W. Joo, S. Y. Moon, S. Singh and. J. H. Parkin 2017 proposed a system called S-detector [8], which describes an enhanced security model developed by authors for detecting smishing attacks for mobile computing. Naive Bayes classifier has been used for the smishing attack detection in smart devices part. This model has the ability to tell the difference between a regular text message and a Smishing message. And the statistical learning method is primarily used to filter. As a result, a text message may be analyzed and SMS phishing can be efficiently detected.

Feature-Based Approach for Detection of Smishing Messages in the Mobile Environment by A. K. Jain and B. B. Gupta in 2019 [9] proposed a method to detect phishing attack messages using a feature-based approach. This method has ten unique characteristics that detect false messages from real messages. These features were implemented on a benchmark database in this study and the performance of the proposed approach was assessed using various classification algorithms. According to the research results, the proposed approach has a true positive ratio of 94.20 percent and a smile message of 98.74 percent overall accuracy. Furthermore, the proposed method is highly effective in detecting zero-hour attacks.

In 2020, G. Sonowal on Detecting Phishing SMS Based on Multiple Correlation Algorithms. [10], states that because of the large number of features in Corpus, researchers have developed various anti-phishing methods and use correlation algorithms to investigate the relevance of features. As a result, this article uses a machine-learning method to compare four rank correlation algorithms, including Pearson rank correlation, Spearman's rank correlation, Kendall rank correlation, and point biserial rank correlation, to discover the optimum features set for detecting Smishing messages. The analysis found that the AdaBoost classifier had a higher level of accuracy. Further investigation reveals that the classifier with the ranking algorithm, Kendall rank correlation, outperformed the other correlation algorithms in terms of accuracy. According to the results of this experiment, the ranking algorithm was able to minimize the dimension of features by 61.53 percent while maintaining a 98.40 percent accuracy.

V. R. Hawanna & V. Y. Kulkarni. in 2016 did research based on the topic A Novel Algorithm to Detect Phishing URLs. [11] It gives a suggestion for an algorithm based on heuristic anti-phishing technical approaches to determine whether a URL is phishing or not in this work. This technique can be used to detect phishing URLs that are both old and new. In the case of known Phishing URLs, it can provide a very quick reaction. If the supplied URL is suspected of being phishing, this algorithm will display an alarm message. Otherwise, it will display a safe message. This algorithm allows users to acquire phishing results fast and easily by directly entering the URL as an input. This solution only works with HTTP URLs.

In 2018, Sonowal, G. & Kuppusamy suggested a model based on machine learning algorithms named "SmiDCA" for detecting smishing communications (Sonowal & Kuppusamy, 2018). Authors selected to extract the 39 most significant elements from smishing messages using correlation methods in their model. After that, they used four machine learning classifiers to assess their model's performance. Random Forest, Decision Tree, Support Vector Machine, and

AdaBoost were the four classifiers. With Random Forest Classifier, the experimental evaluation of this model revealed a 96.4 percent accuracy.

III. Discussion

A. Analysis of Survey results

According to the above studies, there is a high probability that people will be affected by mobile phishing attacks. Therefore, a survey was conducted to get an idea of the awareness in the society about mobile phishing attacks. By using a group of 30 people between the ages of 20 and 50. The reason to choose that age limit is many people in that age limit are using smartphones nowadays. And also, they are using popular social media apps like Facebook, Messenger, and WhatsApp. Also, many people in that age limit regularly check their messages, emails to get in touch with updates in the world. So, their probability of getting affected by phishing attacks is high.

First, it was checked whether the selected people know about phishing attacks. For that question, 63.8% of people answered saying no. Only 36.7% of people have an idea about phishing attacks. Next, it was questioned to check whether they have received a message that pretended to be a mobile phishing attack. To those who don't have an idea about phishing attacks, it was provided an example message to identify. For this, 90% of people said that they have got a message like that at least once. Only 10% of people said they didn't receive that kind of message ever. From those who said yes, the survey asked about the media that they got that message. 93.3% of people answered as a text message, 56.7% answered by saying WhatsApp, 10% answered saying Instagram, 46.7% answered by saying Facebook messenger and 16.7% answered by saying email. So, the majority got mobile phishing attacks through text messages.

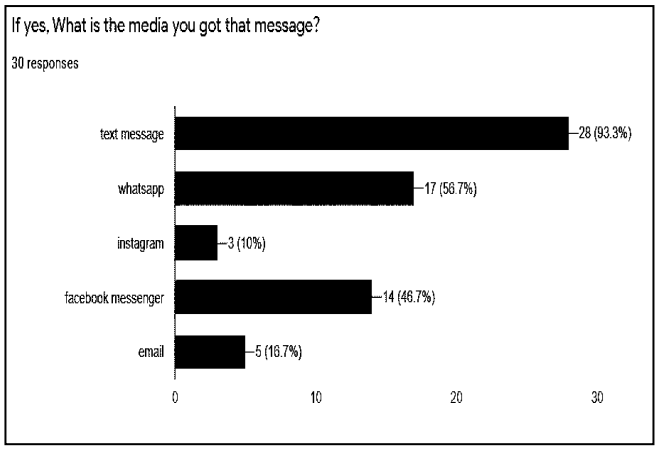

Figure 1: Mobile phishing attacks mostly coming ways

According to the data gathered from the survey, 90% of people have clicked a link in those phishing attack messages. Moreover, 66.7% of people have provided their private details which are asked through that link. Because of this thing they may get into trouble by proving private details to unauthorized parties. Then it was asked whether they have used a system or option which has the ability to identify a harmful phishing attack. 80% of the majority of people said no. Then finally, the survey asked whether it is useful and important to them if there is a system to detect and notify mobile phishing attacks. So, the majority answered yes and it is 93.3%.

Question	Yes	No
Do you have an idea about phishing attack?	36.7% (11)	63.3% (19)
Have you ever received a message that seems to be an phishing attack? (Ex: Congratulations!! You won a brand new iphone. To purchase it go to the below link and provide your details. <link>)	90% (27)	10% (3)
Have you ever clicked that kind of messages?	90% (27)	10% (3)
Have you ever provided your private details to them?	66.7% (20)	33.3% (10)
Have you ever used a system or option to detect phishing attacks that prevents you by clicking those harmful links?	80% (24)	20% (6)
Do you think it's useful and important to have a system to detect and notify mobile phishing attacks?	93.3% (28)	6.7% (2)

Table 2: Survey questions and results

So according to information gathered through this survey, there is a huge need for a system to detect and notify mobile phishing attacks.

B. Analysis of the findings in the Literature Review

Table 2 shows the features applied in systems and approaches which were discussed in the literature review part. Existing research work has been done against machine learning algorithms, Random Forest algorithm, Feature-based technique, Optical Character Recognition, Tag & APK check, Rule-based approach, Naïve Bayes classifier, Heuristic approach, Support Vector Machine, Rank correlation algorithms, Decision tree, and Ada boost classifier.

Machine learning algorithms were used in many approaches. D. Soni and S. Mishra proposed a method using machine learning algorithms to detect mobile phishing attacks. Machine learning algorithms are used for the message classification part. Also, that approach includes tag check and APK check. Another approach was proposed by G. Sonowal which has machine learning algorithms along with correlational algorithms. A system named "SmiDCA" also includes machine learning algorithms with Random Forest, Decision tree algorithms, and ADA boost classifier. This Ada boost classifier is in another system along with the Heuristic approach proposed by Hawanna & Kulkarni. According to findings, another important classifier is the Ada boost Classifier. As the researchers show the AdaBoost classifier had a higher level of accuracy. Jain & Gupta implemented a system using rule-based algorithms. This classification algorithm was used to train nine outstanding rules which can identify mobile phishing attacks easily. J. W. Joo, S. Y. Moon, S. Singh and. J. H. Park proposed an approach with Naïve Bayes Classifier which has been used for the smishing attack detection in smart devices part. This model has the ability to tell the difference between a regular text message and a Smishing message.

Name of the paper	Tag & APK check	Machine learning algorithms	Rule based approach	Feature based technique	Optical Character Recognition	Naïve Bayes classifier	Heuristic approach	Random Forest	Rank correlation algorithms	Decision tree	Support Vector Machine	Ada boost classifier
(Soni & Mishra, 2019)	✓	✓	×	×	×	×	×	×	×	×	×	×
(Jain & Gupta, 2017)	×	×	✓	×	×	×	×	×	×	×	×	×
(Wu, et al., 2014)	×	×	×	×	✓	×	×	×	×	×	×	×
(Joo, et al., 2017)	×	×	×	×	×	✓	×	×	×	×	×	×
(Jain & Gupta, 2019)	×	×	×	✓	×	×	×	✓	×	×	×	×
(Sonowal, 2020)	×	✓	×	×	×	×	×	×	✓	×	×	×
(Hawanna & Kulkarni, 2016), 2016	×	×	×	×	×	×	✓	×	×	×	×	✓
SmiDCA (Sonowal & Kuppusamy, 2018)	×	✓	×	×	×	×	×	✓	×	✓	✓	✓

Table 2: Algorithms and techniques used in each approach

Based on this analysis it can say that machine learning algorithms, random forest algorithms, and Ada boost classifiers are the techniques that were widely used in detecting smishing attacks. Moreover, these techniques have the ability to give results with more accuracy.

IV. Conclusion

Smartphones have become an attractive target for smishers due to a lack of information among mobile users regarding security measures and the users' less secure behavior. Smisher is a hacker who sends a text message to a victim with malicious URLs inserted in it. Smishing, often known as text-based phishing, is a common threat that targets mobile devices. Smishing attacks attempt to obtain private information from users by misleading and tricking them. This paper contains an overview of studies conducted by several researchers in order to detect mobile phishing attacks.

This research work includes the results of a survey conducted among individuals in society to raise awareness about mobile phishing attacks and to identify the need for a mobile phishing detection system. According to survey analysis, there is a huge need for such a system in society today. According to the review, the main identified phishing attack techniques are S-detector, Mobi-Fish, SmiDCA, etc. For detecting smishing attacks, researchers have used various approaches and algorithms. Widely used techniques and algorithms are machine learning algorithms, Random Forest algorithm, Feature-based technique, Optical Character Recognition, Tag & APK check, Rule-based approach, Naïve Bayes classifier, Heuristic approach, Support Vector Machine, Rank correlation algorithms, Decision tree, and Ada boost classifier. Some systems have been made by combining two or three algorithms also. If an approach can produce a result with high accuracy it can identify as an approach with high accuracy.

Acknowledgment

This research was greatly supported by General Sir John Kotelawala Defence University and I would like to pay my gratitude to the lecturers of the Faculty of Computing for their guidance.

References

[1] D. Soni and S. Mishra, "A Content-Based Approach for detecting Smishing in Mobile Environment.," in *SUSCOM-2019*, 2019.

[2] A. K. Jain and . B. B. Gupta, "Rule based framework for Detection of Smishing Messages in Mobile Environment.," in *6th International Conference on Smart Computing and Communication, ICSCC 2017*, Kurukshetra, India, 2017.

[3] L. Wu, X. Du and J. Wu, "MobiFish: A lightweight anti-phishing scheme for

mobile phones," in *2014 23rd International Conference on Computer Communication and Networks (ICCCN)*, 2014.

[4] J. W. Joo, S. Y. Moon, S. Singh and . J. H. Park, "S-Detector: an enhanced security model for detecting Smishing attack for mobile computing," *elecommun System*, pp. 29-38, 2017.

[5] A. K. Jain and B. B. Gupta, "Feature Based Approach for Detection of Smishing Messages in the Mobile Environment," *Journal of Information Technology Research* , pp. 17-25, 2019.

[6] G. Sonowal, "Detecting Phishing SMS Based on Multiple Correlation Algorithms," no. doi: 10.1007/s42979-020-00377-8, 2020.

[7] . V. R. Hawanna and . V. Y. Kulkarni, "A Novel Algorithm to Detect Phishing URLs-Varsharani Ramdas Hawanna," in *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) International Institute of Information Technology (I²IT),* , Pune, 2016.

[8] G. Sonowal and K. S. Kuppusamy, "SmiDCA: An Anti-Smishing Model with Machine Learning Approach," *The Computer Journal*, vol. 61, no. 8, pp. 1143-1157, 2018.

[9] M. Rosenthal, "Tessian," 16 September 2021. [Online]. Available: https://www.tessian.com/blog/phishing-statistics-2020/ .

[10] B. Martens, "SafelyDetectives," 2021. [Online]. Available: 11 Facts + Stats on Smishing (SMS Phishing) in 2021 (safetydetectives.com).

[11] M. LAUDON, "Proofpoint," 02 November 2020. [Online]. Available: https://www.proofpoint.com/us/blog/threat-protection/mobile-phishing-increases-more-300-2020-chaos-continues.

[12] "nextCaller," 08 April 2020. [Online]. Available: https://nextcaller.com/blog/next-caller-covid-19-fraud-report/ .