A Mobile Anti-Phishing System Using Linkguard Algorithm

*Elizabeth A. Amusan, *Oluwinka T. Adedeji, *Oluwaseun Alade, 3Funmilola A. Ajala and 3Kayode Ibidapo

1Department of Cyber Security Science 2Department of Information Systems 3Department of Computer Science
Ladoke Akintola University of Technology, Ogbomoso, Nigeria

Received: 17-JUN-2021; Reviewed: 17-AUG-2021; Accepted: 25-AUG-2021

http://dx.doi.org/10.46792/fuoyejet.v6i3.666

Abstract- Phishing is a type of cyber-attack where the attacker deploys a combination of social engineering and technological skills to trick users into revealing private information like bank account details, usernames and passwords by creating an imitation of an existing web page. This research developed an Android-based anti-phishing system leveraging on the ubiquity of mobile devices and their increasing applications for business and personal purposes. The system was developed by implementing an end-host based algorithm called the Linkguard algorithm which is effective in detecting and preventing both known and unknown phishing attacks. A total of over 500 links which was a blend of both phishing and non-phishing links were collected from Phishtank and Alexa to validate the proposed system which achieved an accuracy of 96%. It is recommended that concerted efforts should be geared towards continuously sensitizing users to improve their phishing awareness as this cybercrime keeps evolving and users play a very crucial role in mitigating it.

Keywords- android, cybercrime, Linkguard algorithm, mobile, phishing

1 INTRODUCTION

Phishing is a cybercrime launched by cybercriminals disguising as genuine individuals in an illegal attempt to get sensitive information from users by using spoofed emails and fake websites. The cyber attackers use a combination of social engineering and technological skills to lure users into revealing private information like bank account details, usernames, passwords and such like (APWG, 2018; Sarannia and Padma, 2014). Unsuspectingely, users or victims give out these pieces of information themselves by clicking on links that appear familiar or similar to that of a trusted contact. Consequently, victims of phishing attacks which could be individuals or corporate organizations are at risk of financial loss as well as loss of confidentiality of valuable information (Joseph and Singh, 2018).

The pervasiveness of mobile devices for business or personal use on the internet and the exponential rise in the number of malware-based campaigns have expanded mobile cyber-threats (Bottazzi et al, 2015). These mobile devices are basically driven by Android and iOS as the two principal operating systems with 99.19% market share distributed as 72.72% for Android and 26.47% for iOS (StatCounter, 2021). While researchers have proposed various approaches to combat this cybercrime, most anti-phishing techniques exist as solutions in desktop environments (Sarannia and Padma, 2014; Joseph and Singh, 2018; Akinwode, Adelakun and Olatunde, 2018; Mohith et al, 2020; Dhanawde, Dhanie and Waghmare, 2020). From literature, very few studies resulting in tools supporting mobility have been reported despite the enormity of activities carried out with these mobile devices (Amro, 2018; Shahriar, Klintic and Clincy (2015)).

Furthermore, several machine learning-based approaches designed and implemented to mitigate phishing are difficult to use by the average non-technical user (Mohith et al, 2020; Sahingo et al, 2019). This is worsened still with the fact that such algorithms are computationally complex and resource-intensive, thereby limiting their support of mobility.

This paper therefore seeks to develop an Android-based phishing detection system through the implementation of Linkguard algorithm which is an efficient and lightweight end-host based algorithm capable of detecting both known and unknown phishing links.

The rest of this paper is structured as follows: section 2 presents the literature review followed by methodology in section 3. Implementation as well as evaluation results are presented in section 4 while conclusion and future work are covered in section 5.

2 LITERATURE REVIEW

Studies have proven that there is almost no difference between the genuine links and the phishing hyperlinks, however, under listed here are some characteristics that the phishing hyperlinks share in common (Naresh, Sagar and Reddy, 2013; Jain and Gupta, 2017; Gupta, Singh and Jayant, 2019; Joseph and Singh, 2018):

i. The visual (displayed) link and the actual link are not the same.

ii. The impostors or cybercriminals regularly use fake DNS names that appear similar to that of a trusted contact.

iii. Special tricks are used to maliciously encode the hyperlinks.

iv. The impostors regularly use fake DNS names that appears like that of the target site though not exactly the same.

2.1 CLASSIFICATION OF PHISHING ATTACKS

The classification of phishing attacks is done on the basis of the strategies deployed by impostors in order to access victims’ information. It typically incorporates social
engineering and technical subterfuge (Gupta, Singh and Jayant, 2019).

**Social Engineering-Based Attacks:** Cyber-attacks in this category leverage mental and emotional manipulation of people thus luring them into completing some actions that will consequently disclose their private information. Social engineering based phishing attacks are further classified as: spoofed email phishing and fake websites. Furthermore, spoofed email phishing could be spear phishing, if the attack is aimed at a particular person or group of persons or whaling if aimed at a high-ranking person or employee (Almomani, Gupta and Wan, 2013). Fake websites that have similar visual appearance like a genuine one are equally used to phish a user.

**Technical Subterfuge:** Phishing attacks here are perpetrated in such a way that the cybercriminal sends malicious codes alongside emails or websites. It could be any of: cross-site scripting (XSS), malware phishing, DNS poisoning and session hijacking (Almomani et al, 2013).

### 2.2 Phishing Detection Approaches

Different anti-phishing approaches have been proposed, these approaches are content-based techniques, black lists, and white lists, visual similarity and URL based. In content-based techniques, the content of the website is used to determine whether it is a phishing site or not (Mao et al, 2017). While black lists and white lists compare the requested URL against either a black list of phishing URLs, or a white list of non-phishing ones (Memon and Khan, 2013). Visual similarity approach determines similarity features like text pieces, images and the overall visual appearance between two web pages and calculates this similarity by comparing the features thereby computing a similarity score (Dhanawde, Dhanje and Waghmare, 2020). URL detection are often carried out using common detection algorithms like Random forest, Support Vector Machine, Decision trees, etc. where length of URL, number of special characters, IP address are the commonly used features.

Shahriar, Klintic and Clincy (2015) provided a mobile anti-phishing taxonomy that highlights all possible phishing attacks and corresponding mitigation techniques. The authors stressed the dearth of solutions developed specifically for mobile devices apart from the black and white lists.

In Bottazzi et al. (2015), authors proposed a framework for the detection of phishing in mobile devices using machine learning algorithms as classifiers whose performances were evaluated in terms of precision, accuracy, true positive rate and false positive rate. However, in an attempt not to overload the computational capability of the mobile devices, the classifiers were built outside the mobile device which makes it challenging for constant update of the model.

Arising from the review of related works, most of the mobile anti-phishing solutions do not perform their detection directly on the mobile devices, rather, they employ a server module to carry out the detection remotely. This is observed as a trade-off in terms of performance, scalability and constant update hence this work therefore proposed a phishing detection system that is mobile and simple to use through the implementation of Linkguard algorithm which is an efficient and lightweight end-host based algorithm capable of detecting both known and unknown phishing attacks.

### 3 Methodology

#### 3.1 Linkguard Algorithm

This algorithm is efficient in detecting and preventing both known and unknown phishing links. It operates by analysing the difference between the actual links and the visual (presented or displayed) link, thus calculating the similarities of URI (Uniform Resource Identifier) (Sarannia and Padma, 2014). Figure 1 shows the composition of the Linkguard algorithm while Figure 2 presents the actual steps:

i. Communication: this component collects data associated with user input, as well as send the information to the analyzer.

ii. Database: this part is responsible for the storage of the whitelist, blacklist, as well as the URL supplied by the user.

iii. Analyzer: it is a very important component as it is responsible for the analysis of data supplied by the communication and database, it also sends the corresponding output to the alert component and logger.

iv. Alert: it alerts the user as soon as it receives a warning message from the analyzer.

v. Logger: it serves to archive related information for reference purpose.

![Fig. 1: Architecture of the Linkguard Algorithm (adapted from Sarannia and Padma, 2014)](image-url)

### 3.2 Model Design of The Proposed System

The purpose of this research is to design an open-source android-based system that is capable of detecting phishing URLs using the operations of the Linkguard algorithm by classifying links as either phishing or non-phishing links. Figure 3 is a flowchart of the proposed anti-phishing system which works by prompting users to input a suspected link in a text button. The system has properties of a web browser embedded in it which categorizes the links based on the result generated from the Google safe browsing archive. The result generated is in form of a visual display on the screen which confirms phishing or otherwise. For the legitimate URLs, the link will load successfully on the mobile app, otherwise, the user will be alerted accordingly.
3.3 ALGORITHM IMPLEMENTATION FOR URL DETECTION

The classification of phishing and non-phishing links is done when the system:
1. Checks for auto login requests from the host.
2. Checks for http response code 401 and 402 which is the most common phishing logic.
3. Check if the websites produce an SSL error. An SSL (secure socket layer) certificate error is when a browser cannot verify the SSL certificate installed on a website which means that website is insecure. It is a standard security technology for encrypting information between a visitor’s browser and your website. It helps to keep sensitive information like passwords and payment information safe. Safe sites can be identified by HTTPS in their URLs and the padlock icon in the address bar.
4. Check if the host of the link is flagged by Google Safe Browsing which is the most important factor. Google Safe Browsing is one of the numerous public API services offered by Google which provides a list of URLs for web resources. These web resources have malwares or phishing contents.

```
intLinkGuard(v_link, a_link) {
  1 v_dns = GetDNSName(v_link);
  2 a_dns = GetDNSName(a_link);
  3 if ((v_dns and a_dns are not empty) and (v_dns != a_dns))
    4 return PHISHING;
    5 return POSSIBLE_PHISHING;
    6 if (a_dns is dotted decimal)
    7 return POSSIBLE_PHISHING;
    8 if (a_dns or v_dns is encoded)
    9 return POSSIBLE_PHISHING;
  10 v_link2 = decode (v_link);
  11 a_link2 = decode (a_link);
  12 return LinkGuard(v_link2, a_link2);
  13 */
  14 /* analyze the domain name for possible phishing */
  15 if(v_dns is NULL)
    16 return AnalyzeDNS(a_link); }
  17 if (actual_dns in blacklist)
    18 return PHISHING;
  19 if (actual_dns in whitelist)
    20 return NOTPHISHING;
  21 for (each item prev_dns in seed_set)
    22 bv = Similarity(prev_dns, actual_link);
    23 if (bv == true)
    24 return POSSIBLE_PHISHING;
    25 return NO_PHISHING; }
float Similarity (str, actual_link) {
  32 if (str is part of actual_link)
    33 return true;
    34 intmaxlen = the maximum string lengths of str and actual_dns;
    35 intminchange = the minimum number of changes needed to transform str to actual_dns (or vice verse);
    36 if (thresh<(maxlen-minchange)/maxlen<1)
    37 return true
    38 return false; }
```

The features used for the extraction of links are: Suspicious char, http, Host flagged by Google safe browsing, Prefix Suffix, SSL error, Auto login request and http response code 401 and 402.

3.4 IMPLEMENTATION ENVIRONMENT

The implementation environment for the development of the mobile anti-phishing system include: an Android operating system (Android 5.0 lollipop and above), any android phone of 512MB RAM upwards and with 8GB internal storage.

4 RESULTS AND DISCUSSION

A prototype of the android-based phishing detection system in form of an Android-based application, KayPhish was developed. To verify the effectiveness of KayPhish anti-phishing system for the link detection section, 500 phishing links were collected from PhishTank and Alexa archive and tested based on search results returned by Google safe browsing. Presented in this section are some snapshots of the developed system shown in Figures 4-7.
Figure 4 represents the home page of the link scanner where a user is allowed to supply the URL of the link to be scanned. When a link has been supplied, it is then highlighted for scanning as shown in Figure 5. The outcome of the scanning would result in classifying the link as either phishing or non-phishing as displayed in Figures 6 and 7 respectively.

4.1 SYSTEM EVALUATION

Because the Linkguard algorithm is a rule-based heuristic algorithm, its performance was evaluated in terms of accuracy and precision which are derivatives of True Positives, False Positives, False Negatives and True Negatives. True Positive (TP) indicates that a phishing URL or link is correctly classified as phishing. False Positive (FP) indicates that a non-phishing URL is incorrectly classified as phishing. False Negative (FN) indicates that a URL is indeed a phishing link but incorrectly classified as legitimate. True Negative (TN) indicates that a non-phishing URL is classified correctly as non-phishing.

Precision indicates the ratio between True Positives and the total number of links tested.

\[ \text{Precision} = \frac{TP}{TP + FP} \]  

Accuracy is the measure of overall correct classifications given by:

\[ A = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100 \]

Of the 500 links tested, we have:

\[ TP= 250, \ TN=230, \ FN=11 \quad \text{and} \quad FP=09 \]

Table 1. System Evaluation Results

| Metrics   | Results |
|-----------|---------|
| Precision | 0.97    |
| Accuracy  | 96%     |

5 CONCLUSION AND FUTURE WORK

This research has implemented the Linkguard algorithm in the development of an Android-based anti-phishing system due to the expanded cyber-threats towards mobile communications and the increasing spread of Android phones for business and personal use. The use of this system provides security against phishing attacks in real time with simple and friendly user interface coupled with...
warnings when a link is not trusted. The accuracy of the
developed system could be further improved on. Also, it
is recommended that users be continuously educated to
increase their phishing awareness as cyber-attack keeps
evolving and users play a crucial role in mitigating it.

REFERENCES
Akineye R O, Adelakun J A and Olatunde K V. (2018).
Detection And Prevention of Phishing Attack Using Linkguard Algorithm. Journal of Information, 4(1), 10-23 ISSN: 2520-7652 DOI: 10.18488/Journal.104.2018.41.10.23.
Almomani B, Gupta, B. and Wan, T. (2013). Phishing dynamic evolving neural fuzzy framework for online detection “Zero-day” phishing email. Indian Journal of Science and Technology, 6(1), 3960–3964.
Almomani, A., Gupta, B., B., Atawneh, S., Meulenberg, A., & Almomani, E. (2013). A Survey of Phishing Email Filtering Techniques. IEEE Communications Surveys & Tutorials, 15(4), 2070–2090.
Amro, B. (2018): Phishing Techniques in Mobile Devices. Journal of Computer and Communications, 6, 27-35. https://doi.org/10.4236/jcc.2018.62003.
Anti-Phishing Working Group (APWG) (2018): The APWG Phishing Activity Trends Report.
Bottazzi G., Casalicchio E., Cingolani D., Marturana F. and Piu M. (2015): MP-Shield: A Framework for Phishing Detection in Mobile Devices. IEEE International Conference on Computer and Information Technology, Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, 26-28 October 2015, 1977-1983. https://doi.org/10.1109/CIT/ICCC/DASC/PICOM.2015.293.
Dhanawde H., Dhanje P. and Waghmure S. (2020): Phishing Attack Detection and Prevention using Linkguard Algorithm. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 07 Issue: 03. pp. 3947-3951. Available online at: www.irjet.net.
Gupta S., Singh P. and Jayant K.P. (2019). A Comprehensive Study on Phishing Attack Taxonomy and LinkGuard Algorithm. International Journal of Engineering Science and Computing, ISSN: 2321-3361, Volume 9, Issue 7, pp. 23190-23192.
Google Safe Browsing API,” http://code.google.com/apis/safebrowsing/developers/guides.html. Last accessed: 5th May, 2021.
Jain AK, Gupta BB (2017). Phishing Detection: Analysis of Visual Similarity Based Approaches. Secur Commun Netw, 2017:1–20. https://doi.org/10.1155/2017/5421046.
Joseph A and Singh J (2018). Real Time Detection of Phishing Attacks Using a Variant of Linkguard Algorithm. Journal of Computational and Theoretical Nanoscience Vol. 15, pp.3303–3307.
Mao J, Tian W, Li P., Wei T., Liang Z. (2017). Phishing-alarm: Robust and Efficient Phishing Detection via Page Component Similarity. IEEE Access 5,17020–17030. https://doi.org/10.1109/access.2017.2743528.
Memon, I.K. and Khan, M.K. (2013). Anti Phishing for Mid-Range Mobile Phones. International Journal of Computer and Communication Engineering, 2, 115-119.
Mohith HR., Adithya MV., Prasad G and Vinay S. (2020). Development of Anti-phishing Browser Based on Random Forest and Rule of Extraction Framework. Cybersecurity, 3(20), Springer, pp.1-14. https://doi.org/10.1186/s42400-020-00059-1.
Naresh U., Sagar V. and Reddy M.(2013). Intelligent Phishing Website Detection and Prevention System by Using Link Guard Algorithm. IOSR Journal of Computer Engineering, vol. 14, pp. 28-36. Available at www.iosrjournals.org.
Phishtank. https://www.phishtank.com Last accessed: 25th May, 2021
Sahingoz OK, Buber E, Demir O, Dirir B (2019). Machine Learning Based Phishing Detection from URLs. Expert Syst Appl 117:345–357. https://doi.org/10.1016/j.eswa.2018.09.029
Saranya A. and Padma U. (2014). Prevention Model for Phishing Attacks in Web Applications using Linkguard Algorithm. International Journal of Innovative Research in Computer and Communication Engineering, 2(1), pp. 2889-2896.
Shahriar H., Klintic T. and Clincy V. (2015). Mobile Phishing Attacks and Mitigation Techniques. Journal of Information Security, Scientific Research Publishing, 6, pp 206-212. Available online at: http://dx.doi.org/10.4236/jis.2015.63021.
StatCounter (2021). Available at: https://gs.statcounter.com/os-market-share/mobile/worldwide. Last accessed: 15th May, 2021.

© 2021 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti. This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/) http://dx.doi.org/10.46792/foyej.et.613.666 http://journal.engineering.fuoye.edu.ng/