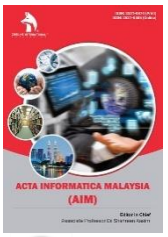
ZIBELINE INTERNATIONAL™ PUBLISHING

ISSN: 2521-0874 (Print)
ISSN: 2521-0505 (Online)
CODEN: AIMCCO

Acta Informatica Malaysia (AIM)

DOI: http://doi.org/10.26480/aim.02.2021.53.57



RESEARCH ARTICLE

AN INCREMENTAL COMPONENT-BASED SYSTEM FOR MITIGATING PHISHING ATTACKS

A.A. Orunsolu[a], S.O Kareem[a], S.I Salawu[a], K.S Famuyiwa[b] and O.J Elugbadebo[c]

[a]Department of Computer Science, Moshood Abiola Polytechnic, Abeokuta South-West Nigeria
[b]Department of Computer Science, DS Adegbenro ICT Polytechnic, Itori, South-West Nigeria
[c]Department of Computer Science, Federal College of Education Abeokuta South-West Nigeria
**Corresponding Author Email: orunsolu.abdul@mapoly.edu.ng*

ARTICLE DETAILS

Article History:

Received 08 November 2021
Accepted 11 December 2021
Available online 21 December 2021

ABSTRACT

Heuristics-based anti-phishing systems are the most deployed solutions in cyberspace. This is due to their generalization feature which is capable of preventing zero-day attacks. However, this promising anti-phishing countermeasure is faced with the problem of managing the feature corpus for scale and complexity. In addition, managing the redundant feature during an upgrade is another factor. In this paper, an incremental component system is proposed to this problem. The proposed framework consists of atomic units and composite units. These units provide encapsulations through which feature corpus is built with ease of management and upgrade. The scheme provides both generic and specific frameworks for designing an anti-phishing scheme with effective management of feature vectors. Therefore, maintenance and updates can be integrated without the total collapse of the system. Hence, the proposed framework can be described as being fault tolerant.

KEYWORDS

Incremental component system, Heuristics, Feature vector, software design, Phishing, Scalability

1. INTRODUCTION

The unprecedented reliance on the Internet and the emerging domain of the Internet of Things (IoT) have created a new social world order in which social interaction and business communication are done via networked computers. This new world order has triggered a dramatic rise in malicious activities in which phishing is one of the most severe of such activities faced by Internet users (Varshney et al., 2016a; APWG report 2020; Orunsolu et al., 2019). Phishing has been increasing in intensity and impact over time as the majority of phishing scams are formatted to appear from a legitimate source with sophisticated appeal and tricks. Phishers always take advantage of inherent human nature that generally ignores critical warning messages.

Phishing is a severe online threat that fraudulently acquired sensitive credentials from online users via fake websites or messages. In most instances, phishers used the acquired credentials to commit identity theft on behalf of the victims which often lead to significant damages bordering from financial losses to personal losses (Adebowale et al., 2018; Qabajeh et al., 2018; Mao et al., 2019). This problem is now aggravated with the emerging trend of IoT social network sites, photo-sharing sites, coronavirus pandemics etc. For instance, an FBI report indicated that phishing was the most common type of cybercrime in 2020 with more than 75% of global organizations experiencing one kind of phishing attack or the other mostly arrived via email (Figure 1). In addition, the most impersonated online brands in the First quarter of 2021 such as Microsoft, Google, PayPal, LinkedIn etc. are mostly patronized online services which put the credentials of most online users at risk of attack (Tessian blog, 2021).

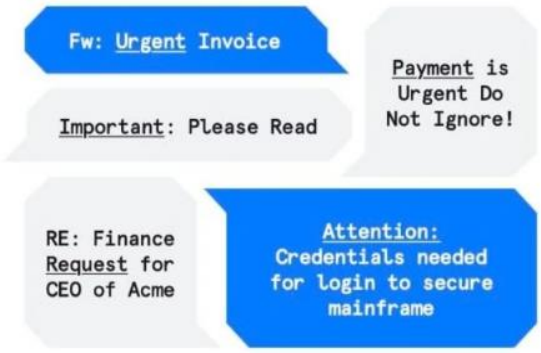


Figure 1: Example of phishing business email (adopted from Tessian blog on phishing)

With the increasing sophistication of phishing attacks, many researchers have designed some methods for mitigating phishing attacks (Han et al., 2012). One of the most common techniques for phishing classification is to use some features to represent a suspicious message (e.g. websites, email or SMS) and apply a learning algorithm to classify the message as phishing or benign based on the extracted features. Several current

Quick Response Code	Access this article online	
	Website: www.actainformaticamalaysia.com	DOI: 10.26480/aim.02.2021.53.57

Cite The Article: A.A. Orunsolu, S.O Kareem, S.I Salawu, K.S Famuyiwa and O.J Elugbadebo (2021). An Incremental Component-Based System For Mitigating Phishing Attacks. *Acta Informatica Malaysia , 5(2): 53-57.*

countermeasures appear to include as many features as can be extracted from phishing instances while identifying a relevant and representative subset of features to construct an accurate classifier remains an interesting issue in the application of machine learning (Chiew et al., 2015; Gupta et al., 2016). In addition, managing scale and complexity in the construction of feature set to meet future changes is another factor. In most cases, heuristics-based anti-phishing schemes are often limited in practical scenarios (e.g. critical online financial transactions) due to maintenance or upgrades and updates which are largely associated with the volumes of features in the discriminative feature corpus (Orunsolu et al., 2019; Adebowale et al., 2018). Hence, effective management of feature corpus is now becoming a major challenge to the anti-phishing community and software vendors.

In this paper, an incremental component-based system is proposed for mitigating phishing attacks through effective feature set management that increase usability, deployment, and low-down time. The key advantage of this scheme is efficient management of scale and complexity in the feature set. This provides for ease of upgrade and removal of redundant features or low relevance features that increase the testing/training time of the system. In this way, maintenance and updates can be integrated without the total collapse of the system. Hence, the proposed framework can be described as being fault tolerant. The main contribution of the scheme is that it provides a generic framework for feature set corpus in any anti-phishing approach. The rest of this paper is as follows: Section 2 discusses the design goals and describe the model of the proposed scheme. Section 3 provides the overview of the incremental component system and section 4 contains the description of the proposed incremental anti-phishing architectural scheme. Section 5 examines an overview of the related works and the paper is concluded in section 6 with insights on future research direction

2. Design Goals and Model Description for The Proposed Acheme

In this design, considerations are given to the following parameters in the construction of the proposed scheme. These parameters are incorporated into the new design to realize an efficient scheme that eliminates some major shortcomings of existing approaches. These parameters are explained as follows:

2.1 Scalable Architecture (SA)

This scheme is designed by viewing the feature set corpus as a class containing units of each feature category from which the set is composed. This provides the merits of incrementally expanding the components or units of the class without affecting the entire system. This is an important attribute of a good anti-phishing approach due to the critical application areas where they are applied.

2.2 Robust Architecture (RA)

The proposed scheme allows for the integration of different feature categories by managing features for relevance or redundancy. This is because the incremental component-based system manages the construction of the feature set for scale and complexity based on the atomic and composite nature of the system's construction.

2.3 Ease of Upgrade (EU)

The Ease of Upgrade is provided in the proposed scheme to give flexibility for including new features or removing obsolete features without disrupting the work of the existing features. This is achieved by adopting the incremental component-based system for building features into the proposed system.

2.4 Model Description for the Proposed scheme

A phishing attack is perpetuated by using the features of online transactions (e.g. website or email or e-chat) fraudulently to obtain users' credentials. Due to poor users' judgement on phishing attacks, enhancement software is designed to classify an online transaction, say *w*, as either phish or legitimate. Equation 1 indicates the simplest scenario that makes phishing possible for online communication.

$$w = \sum_{1}^{n} x_i \quad , n > 0 \tag{1}$$

$$w'' = \{x_1, x_2, \dots x_{n-1}\} \tag{2}$$

$$w''' = \{x_1, x_2, \dots x_{n+1}, y\} \tag{3}$$

In equation 2, the phisher launches the attack by removing some of the features of the original website either in the contents or in the URL or phishing using incomplete scrapping of the legitimate page. This attack may be easily detected by experienced online users. However, in equation 3, the attack is launched by using the complete features of the original page with some additions that may be claimed to be updated by the phishers. This attack is often difficult to detect even for experienced online users especially if the users have not logged on to the page for quite some time. Thus, Equations 2 and 3 provide a premise for extracting as many features as possible from a page to detect its phishing status.

In most existing anti-phishing approaches, feature set X is often represented as:

$$X = \{x_1^1, x_2^2, \dots x_n^n\} \tag{4}$$

where x_n^n is a combination of different feature categories of large dimensions. In such a situation, managing the system for feature updates may result in superfluous computation especially for machine learning approaches as the pre-processing operations would have to run on all the datasets. To overcome this problem, the incremental component system provides the capability of modularizing the anti-phishing scheme through the use of component units and composite attributes (Gowtham et al., 2014; Orunsolu et al., 2019). The component units are used to modularize the heuristics in each class that is employed by an anti-phishing scheme e.g. a, b, c etc. On the other hand, the composite attribute provides a flexible algorithm that incrementally "joins" all the classes in a typical anti-phishing framework or scheme. Hence, the composite attribute indicates the aggregate of all the component units from the system i.e. S_0 ⊆S_1⊆S_2…⊆S_C where S_C is the composite unit built from the subsets of other units.

Based on this approach, a typical anti-phishing scheme is defined using component units as:

$$\begin{aligned} x_a &= x_1 + x_2 + \cdots x_n \\ x_b &= x_1 + x_2 + \cdots . x_t \\ \dots \dots \dots & \qquad \dots \dots . \end{aligned} \tag{5}$$

Where *n* and *t* are dimensions of the heuristics in each feature category. This definition encapsulates the number of heuristics in each feature class and provides the basis for efficient management of each feature class without a total collapse of the resultant feature vector. To combine all the feature class or category into a single system, the composite units or attributes is defined as:

$$X = x_a + x_b + \cdots \dots x_n \tag{6}$$

The resultant feature vector derived from (6) can be trained with a classification algorithm to effectively generalize the detection power of the proposed system. For instance, a Support Vector Machine (SVM) can be applied to the generic framework as depicted in equation (6). The classifier assigns label y to each f_(i)∈w, such that the label y is a binary class represented as:

$$y = \begin{cases} 1 & (i.e. phishing) \\ 0 \ otherwise & (i.e. genuine\ page) \end{cases} \tag{7}$$

Thus, it is the composite unit that computationally attached the system to the classification algorithm.

3. Overview of Incremental Component System

The proposed approach for mitigating phishing attacks are organized into a system using the incremental construction of the component-based systems, which consists of atomic and composite units (Lau et al., 2012). The atomic component is a combination of a computation unit and an invocation connector. In this case, each feature category e.g. x_a is regarded as an atomic component of the system. The computation unit is the function computed on the data in the system while the invocation connector is used to communicate with the other atomic or composite components. In this way, computations are invoked by the atomic units without reliance on other computational units. That is, the computation unit encapsulates computations for which the heuristics in each atomic unit are defined (Figure 2).

On the other hand, the composite component is constructed as a resultant computation unit from atomic components using composition connectors which employ several controls such as sequencing, pipe, looping etc. for

achieving efficient system construction. Sequencing and branching are used as composition connectors for multiple components whereas looping is often used as an adaptor for a single component. For sequencing operation, a sequence and a pipe composition connector are often employed in the design flow. If there are n components in a proposed system using this approach, these components are connected by n composite adaptors.

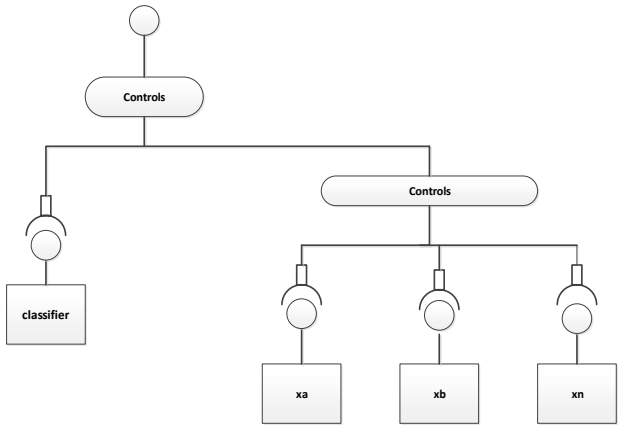


Figure 2: A Generic Framework for Incremental Component System

4. The Proposed Framework

The proposed scheme provides a specific framework for implementing the concept of incremental component systems in mitigating phishing attacks. Figure 3 shows an incremental component system consisting of the login interface, the phishing detector manager and the phishing alert response module. The proposed architecture combines the merits of an incremental component based system to build an efficient anti-phishing scheme. The proposed system provides a generic framework for managing the complexity of the feature set corpus more efficiently. Hence, the superior efficiency of this system stems from the efficient construction of feature class as a unit of a composite system, support integration of historical features without disrupting the system and provide richer configurability and reconfigurability of anti-phishing services with negligible down-time.

The login interface is the first module where a request (e.g. login into a mail server, website) attempting to connect to the webserver is examined. The examination is based on the generation of the Document Object Model (DOM) tree of the requesting page from the web browser. This is because the DOM presents the web document as nodes and object that allows structural representation from a single consistent API. The login interface employs the JSOUP parser to scrape and transverse the DOM of any web document in the proposed approach. The JSOUP library is equipped with API to manipulate data from a web's URL or HTML tags using the DOM, CSS and Jquery techniques thereby providing an efficient way of extracting useful features from any loading page.

The Phishing Detector Manager is the core component of the proposed system where the concept of the incremental component system is implemented. The manager consists of a Feature Selection Module (FSM) which is used to identify all relevant features required for an efficient phishing system. As the number of extracted features can vary from one to many, a ranking method is used to determine a more discriminative feature category. Based on this, three feature categories consisting of URL, Web document properties and Behaviour of webpage were identified. The three components are organized incrementally to produce a system in Figure 4. Each of the feature categories consists of several heuristics that are defined in each level such that at the initial stage, the feature generator is set to 0. Thereafter, the first feature category is added to it to the constructs feature vector x_a which is then communicated to the composite adaptor.

The composite adaptor concatenates the features returned from the previous operation with the features at their level and construct the resultant feature vector which is passed on to the next adaptor until all the components are exhausted. The last component to be connected is usually the machine learning algorithm on which phishing prediction or label is based.

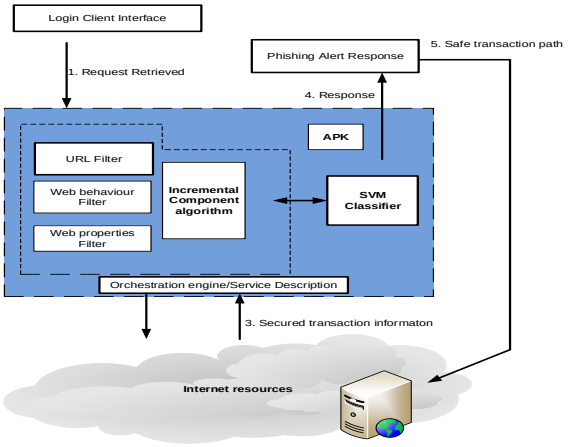


Figure 3: The Incremental Component Approach

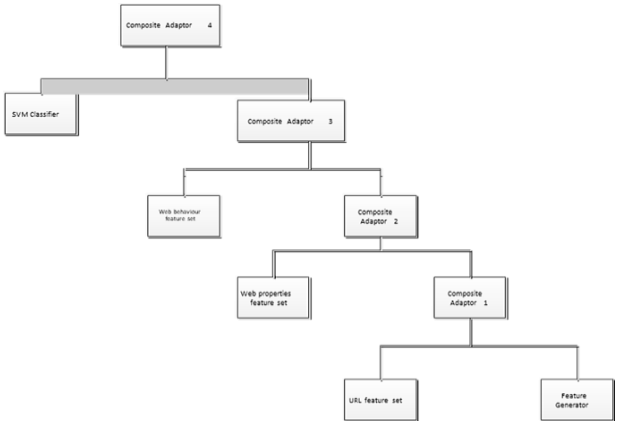


Figure 4: Incremental Based Anti-Phishing Scheme

The incremental component algorithm for the proposed approach is presented as Algorithm 1. The algorithm works by setting the feature generator to 0 and thereafter the first feature category is invoked. In this case, the first feature category is a URL class which consists of several heuristics such as several dots, length of URL, the domain name in URL path etc. These heuristics are semantically defined using the appropriate notation to access them in the DOM structure. The generated feature set is then forwarded to the next feature category. The feature generator is then set to 1 and the resultant feature category is generated when the heuristics of the second feature category is added. The step is repeated until all the feature categories are exhausted. Thereafter, the resultant feature set is normalized, and the classification model is produced to determine the status of the querying page.

Algorithm 1: Incremental Construction Algorithm (ICA)

Input: *Suspicious URL*

Output: *URL status*

Begin

Set feature generator = 0

Invoke the first feature category

Generate feature set for the first feature category

Forward feature set to next feature category

Return ***False***

Else

Feature generator = feature generator + 1

Invoke the second feature category

Generate feature set for the second feature category

Forward resultant feature set to next feature category

Return ***False***

Else

Feature generator = feature generator + 1

Invoke the third feature category

Generate feature set for the third feature category

Forward resultant feature set to next feature category

Return ***False***

Endif

Continue until all feature categories are exhausted

Normalize and vectorize resultant feature vector

Generate the trained classifier based resultant feature corpus

Generate classification model for the classifier

Test the classification model

End

The Phishing Alert Response (PAR) Module provides an appropriate response system for defeating already classified phishing attacks. The PAR computes the appropriateness of the response system by examining the number of factors through which the attack is propagated. For instance, a phishing attack launched through malicious URLs may be easily detected and shut down. However, a phishing attack on a comprised domain or link manipulation within a legitimate may not be easily discovered or known. A Transaction Identification Module is designed as a service within PAR to compute the phishing attack level. The Transaction Identification Module is a service that measures and identifies the threat associated with a classified transaction. With classified transactions, a TIM is proposed to proactively predict the level of seriousness of the attack. The algorithm for TIM is described in Algorithm 2.

Algorithm 2: Transactional Identification Module (TIM)

Input: t_i ***= classified transaction***

Output: ***: ts = threat score, rs = response type***

Start

If $(t_i = null)$ **then** stop

Else if $(t_i = 1 \,.and.\, 0 < t_i < 1\ )$ **then**

Replay (t.history and t.features)

Compute t.impact = Phishing attack level (PAL)

If (PAL = 1) // high level phish i.e. attacks affecting all the features

Deploy HIR // high impact response

IF (1>PAL ≥0 .6) // middle level phish i.e. attacks affecting URL and web features

Deploy MIR // middle impact response

IF (PAL < 0.6) // low level phish i.e attacks affecting URL features only

Deploy LIR // low impact response

End if

End

4.1 High Impact Response

This is a type of response disrupts a sophisticated form of a phishing attack in which the phishers employ strategies that affect as many feature categories as possible. In such attacks, strategies such as injection attacks, HTTPS splitting, obfuscations etc. will have been employed. The high response definition includes using code instructions where malicious HTML codes can be rewritten based on standard web design tags.

4.2 Medium-Impact Response

This is a type of response that accompanied a phishing attack in which some selected feature categories are employed to perpetrate phishing attacks. MIR uses the genuine name of a real website to inform impersonating organizations through an automatic Anti-Phishing Organization Discovery Alert Manager (APODAM). In addition, APODAM can utilize the social networking account to inform the user of such malicious links.

4.3 Low Impact Response

This is a type of response in which a phisher uses low-level tricks in deceiving users. Such attacks are easy to detect if users are informed and can easily be detected by an automated system. Upon detection of such attacks, LIR uses a warning and alert code with a short description of the attack to disrupt the phishing attack.

5. Related Work

Several approaches have been developed by research communities and online security companies to tackle phishing attacks. One of the most popular approaches by the online security community is Google Safe Browsing. However, this approach is browser-specific and may not detect phishing on compromised legitimate domains. Considering the severest of phishing attacks to the safety of online communication, the onus lies on research communities to continuously innovate new approaches to match up with the antics of the cybercriminals. These methods vary from software enhancement approaches to user-awareness programmes.

A group researchers examined the performance assessment of some phishing predictive models based on the concept of minimal feature corpus (Orunsolu et al., 2021). The approach considered several machine learning methods to determine the effectiveness of their computed feature vector in detecting phishing attacks. The approach was experimented with using WEKA and JSoup HTML Parser. Using a 10-fold cross-validation experiment, the results indicated that Random Tree performed better than other classifiers with a minimal feature corpus.

A group researchers proposed a machine learning approach using a Decision Tree and Optimal Features based Artificial Neural Network to detect phishing attacks (Zhu et al., 2020). The approach improved the traditional K-clustering algorithm with incremental selection to remove the duplicate points from the public feature datasets. Then, an optimal feature set was generated using the feature evaluation index, decision tree and local search method. The approach used several features such as domain features, HTML/Javascript features, abnormal behaviour features etc. Experimental analysis indicated the higher performance for the approach. However, the approach does not provide for robust feature management.

A group researchers examined an approach called PhishCalcluator (Orunsolu et al., 2020). The approach adopted URL legitimacy based on a weighting factor to predict the status of a loading URL. The technique achieved significant accuracy without the use of machine learning algorithms. Similarly, Jain and Gupta 2016 examined an auto-updated whitelist technique based on URL and DNS information. The approach was implemented as a client-side scheme for mitigating phishing activities. The approach achieved an accuracy rate of 86.02%. Using search engine technology, Varshney et al. 2016 proposed a phishing detector approach using the domain name and title of a URL. The scheme achieved a remarkable accuracy rate of 99.5%.

A group researchers proposed an active blacklist approach in which new malicious URLs can be effectively predicted from the existing blacklist entries (Prakash et al., 2010). The approach processed blacklisted URLs and produced some variations of the same URL. The method used the IP address, query string substitution, brand name, directory structure similarity and top-level domain replacement to produce the child-URL. The technique produced promising results when applied to real-time analysis of blacklist sites

Afroz and Greenstadt presented a technique called PhishZoo which built profiles of trusted websites based on fuzzy hashing techniques in a whitelisted-based approach on a local database (Afroz and Greenstadt, 2011). The profiles built-in PhishZoo included URL, SSL, images, HTML and scripts of trusted websites. This method checked the cached profile of genuine websites with the querying websites. The approach achieved a significant accuracy rate of about 96% with the possibility of defeating the zero-day attack. However, there is a lack of generalization to new phishing due to human interventions.

An approach that used a whitelist was investigated (Han et al., 2012). The approach called an Automated Individual Whitelist (AIWL) recorded the well-known legitimate URLs visited by online users. The AIWL maintained URLs with their Login User Interface (LUI) information. The LUI is an area where the user inputs his or her details to prevent unhealthy disclosure of confidential information to malicious sites as well as a repository for other information about the visited websites. The URL represents the address of the website. The input area includes the form username path and password path. This method is very effective against first login-attempt and new sites. However, this method can be circumvented if the online users have a low level of internet training.

A group researchers present a dynamic defense approach in which direct and indirect links associated with a malicious page is generated (Gowtham et al., 2014a). The method constructed a target domain set based on a pre-defined Target Identification algorithm to detect the status on a loading page based on the DNS lookup and IP address resolution. However, the

prediction of this approach is largely dependent on the TF-IDF algorithms, search engine speed and DNS lookup. A group researchers investigated an approach called *PhishTackle* where the concept of feature management was given priority (Gowtham et al., 2014b). The approach used the composite web services by Lau et al. 2008 to provide a robust feature management algorithm. The experimental results indicated an accuracy of 99% with a low false-positive rate.

A studied frequency assessment of existing feature vectors to produce a predictive model based on Naïve Bayes and SVM (Orunsolu et al., 2019). The approach provides the first specific incremental component system in phishing detection. The system was evaluated using several experiments that achieved an accuracy of 99.96% with low false positives. Similarly, proposed an anti-phishing scheme using an SVM classifier (Mao et al., 2019). The classifier was trained based on several visual features. Their evaluation produced an accuracy of more than 93.0%. Zouina and Outtaj investigated an SVM classifier based on several URLs in a lightweight phishing detection approach (Zouina and Outtaj, 2017). The authors extracted six features from the URL of a querying page and achieved an accuracy rate of 95.80%. a group researchers investigated a Remove-Replace Feature selection approach (RRFST) from the existing feature corpus with an accuracy of 99.27% on an ensemble of C4.5 and CART (Hota et al., 2018). A group researcher proposed a machine-based approach using a stacking scheme on 20 features obtained from the URL and HTML (Li et al., 2019).

The features were trained using Gradient Boosting Decision Tree, XGBoost and LightGBM in a stacking manner. Their evaluation produced an accuracy of 98.60% accuracy and a 1.54% false alarm rate. A group researchers examined an Adaptive Neuro-Fuzzy Inference approach on several integrated features extracted from the URL, HTML, visual similarity, and logo property of a website (Adebowale et al., 2018). The approach considered Information Gain as criteria for inclusion of features and the method produced an evaluation result of 98.3% accuracy. Summarily, these methods have the advantage of remarkable detection accuracy and low false positives, the training requirement of these classifiers is a limiting factor in a smart environment where superfluous computation associated with such training may produce a significant delay in data gathering, sharing and communication activities of aggregated devices.

6. Conclusions

This article presents an incremental component-based anti-phishing scheme for mitigating phishing attacks. The proposed scheme is novel for providing a generic framework for managing complexity and scalability in a large feature vector corpus. This is usually a challenge especially when more than one feature category is employed in mitigating phishing attacks. The superior attraction of the approach stems from the capability to manage redundant features or new features without disrupting the entire system. The approach is demonstrated by using a three-feature category incremental based approach. In the future, we intend to increase the number of feature categories to examine the contributory effect of large feature categories. In addition, we intend to examine the results with other classifiers such as ANN, NB and other prominent classifiers that are popular in phishing detection.

References

Adebowale, M., Lwin, K., Sanchez, E., and Hossain, M., 2018. Intelligent Web-Phishing Detection and Protection Scheme using integrated Features of Images, Frames and Text. Expert System with Applications.

Chiew, L., Chang, H., Sze, N., and Tiong, K., 2015. Utilization of website logo for phishing detection. Computer and Security Journal.

FBI Report. Accessed 2020.

Gowtham, R., Krishnamurthi, I., 2014a. PhishTackle—a web services architecture for anti-phishing. Cluster Comput

Gowtham, R., Krishnamurthi, I., 2014b. A comprehensive and efficacious architecture for detecting phishing webpages. Journal of Computers and Security Journal.

Hamid, A., and Abawajy, J., 2014. An approach to profiling phishing activities. Journal of computer and security. Elsevier Press.

Han, W., Cao, Y., Bertino, E., and Yong, J., 2012. Using automated individual white-list to protect web digital identities. Expert Systems with Applications.

Hota, H.S., Shrivas, A.K., and Hota, R., 2018. An Ensemble Model for Detecting Phishing Attack with Proposed Remove-Replace Feature Selection Technique. International Conference on Computational Intelligence and Data Science. Procedia Computer Science, Vol. 123, Pp. 900-907

Jain, A., and Gupta, B., 2017. Two-level authentication approach to protect from phishing attacks in real-time. J. Ambient Intell Human Comp., DOI 10.1007/s12652-017-0616-z

Jain, A.K., Gupta, B.B., 2016. A novel approach to protect against phishing attacks at client side using auto-updated white-list. EURASIP J Inf Secur., Pp. 1–11.

Mao, J., Bian, J., Tian, W., Zhu, S., Wei, T., Li. A., and Liang, Z., 2019. Phishing Page detection via classifier from page layout feature. EURASIP Journal of Wireless Communication and Networking, Vol. 43.

Nirmal, K., and Kumar, R., 2020. Analyzing and eliminating phishing threats in IoT, networks and other web applications using the iterative intersection. Journal of Networking and Applications.

Orunsolu, A., Sodiya, A, Kareem, S., and Oladimeji, G., 2021. Performance Assessment of some Phishing predictive models based on Minimal Feature corpus. Journal of Digital Forensics, Security and Law, Vol. 16 (5).

Orunsolu, A., Sodiya, A., and Kareem, S., 2020. LinkCalculator- An Efficient Link-Based Phishing Detection Tool. Acta Informatica Malaysia.

Orunsolu, A., Sodiya, S., and Akinwale, A., 2019. A Predictive Model for Phishing Detection. Journal of King Saud University-Computer and Information Sciences.

Pavithran, D., Shaalan, K., Karaki, J., and Gawanmeh. 2020. Towards building a blockchain framework for IoT. Journal of Cluster computing.

Prakash, P., Kumar, M., Kompella, R., and Gupta, M., 2010. PhishNet: predictive blacklisting to detect phishing attacks Proceedings of 29the Conference on Information Communications.

Qabajeh, I., Thabtah, F., and Chiclana, F., 2018. A recent review of conventional vs. automated cybersecurity anti-phishing techniques. Computer Science Review.

Tan, C., Chiew, L., and Sze, N., 2017. Phishing Webpage Detection Using Weighted URL Tokens for Identity Keywords Retrieval. Lecture Notes in Electrical Engineering. Vol. 398.

Tessian blog report. Accessed 2021

Varshney, G., Misra, M., and Atrey, K., 2016. A phish detector using lightweight search features. Comput Secur., 62, Pp. 213–28.

Zhu, E., Ju, Y., Chen, Z., Liu, F., and Fang, X., 2020. DTOF-ANN: An Artificial Neural Network phishing detection model based on Decision Tree and Optimal Features. Applied Soft Computing Vol. 95.