IST-Africa
IST-Africa 2019 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2019
ISBN: 978-1-905824-62-5

Literal and Conceptual Consistency for Anti-Phishing Solution

Nureni Ayofe AZEEZ, Balikis SALAUDEEN, Charles Van der VYVER
School of Computer Science and Information Systems,
North-West University, Vaal Triangle Campus, South Africa
Email: nurayhn1@gmail.com, Charles.VanDerVyver@nwu.ac.za

Abstract: Phishing is a fraudulent attempt by cybercriminals where the target audience is consulted by either a text message, phone call or e-mail, requesting a classified and sensitive information relating to his financial transaction after presenting himself as a legitimate agent to a financial institution. When the target finally falls into a phishing attack tactics, there will be a terrible financial loss and identity theft. The target who might not initially realize his source and current state of attack can reveal all the confidential details such as credit card digit, online shopping account password, bank verification digits and even employment details. To shield clients from phishing assaults, numerous anti-phishing models have been proposed. It has been established that some techniques that are currently being employed over the years to handle these challenges are not sufficient and capable enough to do the needful in getting the lasting solution. To find a solution to this issue, this work aims at identifying phishing sites in order to guard internet users from being vulnerable to any form of phishing attacks by verifying both the conceptual and literal consistencies between the Uniform Resource Locator (URL) and the web contents. This technique was implemented using java and PHP programming language. The results after the implementation show that the application achieve 98% precision and accuracy; indicating that it is effective in detecting various forms of phishing attacks.
Keywords: Phishing, cybercrime, legitimate, literal and conceptual

1. Introduction

Information and Communication facilities are easy to use and efficient despite the fact that large number of people are using the internet in their daily activities. However, as the internet facilitates convenient access to data and classified and confidential information, such access can also cause an internet user to lose their belongings, specifically, money through the nefarious activities of cybercriminals like hackers and phishers. Phishing, which is no more a new cybercrime, is a strategy being used by criminals to lure and deceive many internet users into revealing their online financial transaction details. Whenever a user is trapped by their deception, the user tends to lose either a big amount of money or confidential information or even both [32]. Any form of phishing attack is usually introduced by sending a link believing to be from a genuine source [36]. The moment a feedback is given by the user, his online details would have been revealed to the phishers. The concept of phishing which came into existence in 1996 has remained a common term in cybersecurity [1].

Apart from e-mail, other platform being used to perpetrate phishing attacks include but not limited to social networking sites, voice messaging, multiplayer games, SMS and instant messaging [2]. The main objective of creating a phishing site is to clandestinely and fraudulently obtain confidential information such as personal identification numbers (PINs), credit card details and passwords [3] and use it nefariously [37].

The attempt to protect online users from this fraudulent financial crime has brought about the establishment of Anti-Phishing Working Group (APWG) in 2013. The group is an international organization that collects phishing information from companies, government agencies, communication companies, law enforcement agencies, etc. affected by phishing attacks from different sources. With the report obtained in 2014 by the Anti-Phishing Working Group (APWG), a total number of 255,000 new attacks are being detected on daily basis [29]. A total of 197,252 phishing cases which shows 18% from previous years was reported the same year. In an effort to safeguard internet users from various attacks and subsequently prevent them from losing money through their online transactions, several anti-phishing solution have been proposed [3]. Some of the approaches are truly good and efficient but not reliable and dependable to cater for the current trend of global phishing strategies [34].

In an effort to detect phishing websites, the authors attempt to check both literal and conceptual consistencies. To the best of our knowledge, this approach seems different both in results and approach as the results obtained justify our position.

1.1 Literal and Conceptual Consistencies

Literal checking of a URL simply means using the contents that the URL itself comprises of, without adding or bringing in other properties and features for test running and verification. This is achieved by following the exact words and theory between the realism and nominalism. The literal and conceptual consistency as used in this work is simply the way the URL and the content of the page are examined. The correlation between them as supposed to what the site itself portrays. Each letter, number and symbol of a URL is divided into various categories to allow for a thorough analysis in order to avoid wrong categorization.

2. Related Work

CANTINA which is a technique for anti-phishing is a unique content-oriented technique to identify and detect various phishing sites by analyzing and examining what is contained in a given webpage for classification as either legitimate or illegitimate [29]. The classification is determined by the Term Frequency -Inverse Document Frequency algorithm (TF-IDF) that uses search engine for extractions and retrieval of information as it does not rely only on surface level characteristics. TF-IDF is an algorithm for retrieving information that can be used to classify and compare documents. It also identifies most weighted words which generates lexical signature [5][6].

Phish Tester Based Approach which was developed by [7] to provide a finite sate machine model which is based on known legitimate and phishing websites behaviours in order to differentiate legitimate and phishing websites in terms of forms submission that is based on random inputs. A set of heuristics combinations was developed to capture the recent up-to-date behaviour of suspicious websites. The technique can assist with information and warning about the likelihood of a phishing sites present in trusted sites and can detect an XSS-based attack which has an edge over some other existing techniques [7].

In an effort to address the problem of phishing, GoldPhish was proposed by Dunlop et. al., 2010. This technique was aimed at identifying and detecting new faces of phishing attacks. It has come to the notice of researchers that phishing sites exist within few days or hours. Because of this development, authors developed GoldPhish to among other things capture the image of the existing website. The image that is captured will be converted to text data which will serve as input and finally be used to retrieve search result. GoldPhish utilizes only the first four results to determine if a site is legitimate [13].

Also, in a similar vein, Garera et. al., 2007 adopted the composition and structure of URLs phishing website identification. They combined numerous heuristics and Google PageRank to decide the status (legitimate or phishing) of a URL [12]. The motive behind this is that any established website will be high-ranked while illegitimate websites will be low-ranked. The results obtained show 1.2% false positive rate and a 4.2% false negative rate [1].

2.1 Anti-Phishing Approach

The presence of phishing has made analysts find different systems for its recognition and diminishing the rate at which clients get bulldozed by it.

a. Blacklist and Whitelist Approaches

Blacklist-based technique keep detailed and up-to-date information about all the phishing websites. Anytime there is information regarding a phishing website, the Internet Protocol address (IP address) where the suspected phishing website is initiated will immediately be included among the blacklisted IP addresses [28].
Anti-phishing platform that specializes in information distribution has been proposed. This system is efficient to authorize both the whitelist and blacklist data. This system which has client side proxy as browser can conveniently authenticate sites by confirming its status without any warning or notice to user [9]. The system is automatic in nature. Anytime, a user enters a URL into the browser, its status will automatically be determined [13]. If the site is found to be a blacklisted one, it will be blocked immediately [33].

b. Heuristic-based Approaches

Heuristic-based approach uses web content and URL signatures to detect phishing behaviors through the extraction of different features like visual similarity. Heuristic can produce true positive and true negative rates and has the ability to detect the moment an attack is launched. The disadvantages of heuristic approach is the fact that it can label a legitimate websites as a phishing website by producing false positive. The approach can improve the precision as well as the accuracy of detection rate. Surprisingly however, there is a high hit rate. Lexical features are used for analyzing various keywords obtained the web content and URL. This is achievable with the aid of heuristic-based approach [14].

c. Visual Similarity-based Approach

Resemblance of layout as well as the general style between likely phishing and legitimate sites is usually calculated and determined by visual similarity [17]. This is obtained majorly by taking into consideration and recognizing a webpage as an entity [15].

d. Pre-Filtering Phase

This phase is the first phase by which possible potential identities and associated domain names will be examined. This phase assists to separate legitimate site from a phishing or suspicious sites and move the latter to the classification stage. Nearly all the legitimate sites have their names registered as their second level domain names. To effectively carry out what this phase is meant for, two things must be put into play: the identity extraction and checking if the consistency matches with the second level domain [29].

e. Identity Extraction

This simply means being able to recognize if a webpage is malicious or legitimate by looking at the URL. Most phishing websites are commonly constructed to confuse their viewer into believing that the URLs are a legitimate one. To identify the possible signs, the first process is to look only at the URL of the page to determine features. This can be viewed properly if the URL is not structured appropriately or a number is used to represent a domain name instead of the usual alphabet. Phishing websites use a URL that looks like a legitimate one, alternatively by putting a legitimate URL among the illegitimate/phishing with the motive of deceiving internet user. After this has been checked and verified and subsequently clear of any suspicion, it can also be checked if it matches a whitelist of a high profile or any other safe site hence the identity extraction is passed [31].

f. Check if Consistency Matches with Second Level Domain

This simply implies that the degree of compatibility of the URL matches with the brand names of the company. The content of the site should be in compatibility with the domain name that appears in the URL of the webpage. To achieve that, some features have to be put into consideration. The page content checks the HTML of the webpage for likely suspicious features and identifies if a site is phishing. The company's logo, images and keywords are also factors to be checked for, but observation shows that most phishing sites copy the company's logo, images and keywords word for word, however this method alone is not considered. Phishing websites are short-lived in nature so they rarely get indexed by search engines [32]. Whenever a click is made by a user, it is usually stored in as search log. The URL that has been frequently and consistently clicked will be classified as legitimate [30].

g. The Classification Phase

Under this phase, there are numerous features to consider in order to discover if a site is phishing or legitimate. This phase emphasizes on: Randomness of URL (RU), Ratio of found domain token (RDT), Position of domain token (CPos), and Conceptual Similarity (CSim).

h. Randomness of URL (RU)

While legitimate sites have precise URL content, the URL contents of a malicious website are usually unrelated to the website and also contain irrelevant long random strings. For example (http://signin.ebay.com.87ab3540af65fa59167f076ea075f9f7.ustsecurity.info/). The long random string in this website (87ab3540af65fa59167f076ea075f9f7) is irrelevant because it does not have any role to play in the website being visited. This might be due to the fact that phishers rapidly generate many malicious sites. There are some features to look out for when considering the randomness of URL. URLs cut across different angles like, IP address, symbols, URL length, HTTPs protocols and so on [30]

Phishers are always in the habit of surreptitiously redirecting the uploaded webpage to suspicious domain. Presently, research has not convincingly shown us that there is a standard length of a URL to differentiate legitimate and phishing sites. However, [17] is of the opinion that any URL with length greater than 54 characters should be considered phishing.

With @ symbol in a URL, the browser disregards all before it and redirects the internet user to the link that appears thereafter. In most cases, phisher adopts this approach to lure and deceive many victims. The user will not know he is being redirected to another page entirely therefore divulge all the important personal information to the site. For example, "http://www.kfc.com@http://www.hacker-site.com" will navigate to the hacker site instead

of kfc. All URLs are being split by reserved symbols like ";", "/", "?", ":", "&", "=", "+" but not with @ and _, These two has a different meaning in the URL and are mostly used by phishers. The domain name needs to be segmented after all the tokens of a URL has been generated.

i. Web-address-makeup.pdf

This represents the last destination, that is, the file that is being looked for. In this case, the file name is pdf (Portable Document Format) File. In some cases, file name could be index.asp, index.htm, index.html etc.

In calculating the randomness of a URL (RU) we adopted the following formula as formulated and used by [27]:

$R(t) = \max (sl\_d(t), sl\_s(t) * \log_2 (as(t))$……...(1)

Each of the parameters is described as follows:
R(t) can simply be defined as randomness of token t.
sl_d(t) is the total number of segments formed by splitting token t into digits.
sl_s(t) is the total number of segments formed by splitting token t into symbols.
as(t) is the total number of letters in token t. the maximal randomness score of the tokens will be the RU of the URL.

j. Position of Domain Token (CPos)

Most phishing sites have lots of sub-domain names than legitimate sites. Research has revealed that the highest number of sub-domain name of any site considered legitimate is 5 while 18 is for phishing site. What is more, there is a standard rule for a given hierarchy of the domain name. The higher the domain tokens, the higher the level of categorization of the webpage. The position of domain token (CPos) can be determined as shown in Algorithm 1 adapted from [27].

Algorithm 1 CPos calculation Algorithm

1: token_list = extract_domainName_token(url)
2: ***if*** *domain is IP* ***then***
3: score = 0
4: ***else***
5: ***for*** *each token in token list do* ***do***
6: allIDX_sum + = index of token
7: ***if*** *token is included in content* ***then***
8: foundIDX_sum + = index of token
9: ***end if***
10: ***end for***
*11: score = foundIDX_sum = (allIDX_sum * number of count of token)*
12: ***end if***
13: return score.

Adapted from [27].

k. Ratio of Found Domain Token (RDT)

Literally, most phishing sites are accessible for just a couple of hours or days hence the URL of phishing sites cannot have a reasonable number of clicks on the internet because of its short existence. It has been established that URLs with frequent and uncountable number of clicks can be regarded as legitimate. Cid list of the identities can be regarded as the second level domain names. The concept entities linked to the entities are other domain tokens obtained from similar data source and detailed in the list of domain token [27]. Legitimate sites should have a greater ratio of tokens at the second level domain that that of any malicious websites [36]. The formula for calculating the ratio of found domain token is given as follows [27]:

$$Score = \left(\sum_{i=1}^{n-1} f_{DT}(t_i) + f_{SLD}(\text{sld}) \right) \ldots\ldots\ldots 2$$

Where,

f_{DT} (t) = 1, t is found in DT list
f_{DT} (t) = 0, t is not found in DT list
f_{SLD} (t) = 1, t is found in SLD list
f_{SLD} (t) = 0, t is not found in SLD list
n = number of domain tokens
sld = second level domain name

Algorithms 2: An Implementation of Anti-Phishing Approach

1. *Enter URL*
2. *Check URL against web content*
   Check if content contains below
   Check for tittle
   Check for link
   Check for styles
   Check if URL has http request: (http, https, :, //, :, [a-z, 0-9]
3. *Check if Cid matches with SLD*
   If passed, check if URL exists in LEGITIMATE database, else insert into LEGITIMATE database
   If not passed, proceed to (4)
4. *Check if URL contains long random irrelevant strings*
   If passed, proceed to (5)
5. *Check if numbers of sub-domain is more than 5*
   If passed, proceed to (6)
6. *Check if URL has a reasonable amount of click records*
   If not passed, proceed to (7)
7. *Check if concept of URL and webpages are similar*
   If passed, check if URL exists in LEGITIMATE database, else INSERT into LEGITIMATE DATABASE.
   If not passed, check if URL exists in PHISHING DATABASE, else insert into PHISHING DATABASE.

3. Implementation and Evaluation of Result

3.1 Legitimate Site

The URL is entered into the query box of the application, the application checks if the URL exists online by checking PhishTank. If it exists in the PhishTank, the URL is considered a phishing one else it goes through all the phases (Pre-Filtering Phase and Classification Phase)

where the Randomness of the URL, ratio of found domain token and the positon of the domain token, thereafter the Conceptual similarities before the URL can be confirmed whether it is legitimate or phishing. The presence of the URL on PhishTank database indicates it is a phishing site.

3.2 Phishing Site

When a URL is entered into PhishDetect application, the application runs through the submitted URL, if the URL does not have the features of a legitimate site, the application classifies it as a phishing site. Once the URL is submitted, the application uses all the features embedded inside it to check if it's malicious or legitimate. Once it's confirmed malicious, the application displays all the features checked to tell the user it is a pure phishing website.

3.3 Invalid Site

If the submitted URL does not have any of the attributes of phishing and legitimate or the URL is incomplete - or does not exist online, they are classified by the application as invalid and stored in the invalid database. Figure 3 shows a URL that does not exist online.

3.4 Graphical Representation of Phishdetect Outputs

Figure 4 shows that when the total number of tested URLs was 50, the application returned 42 URLs as legitimate which are true negative, 6 URLs as phishing that is true positive and 2 invalid which means the application could not classify the URL as either legitimate or malicious.

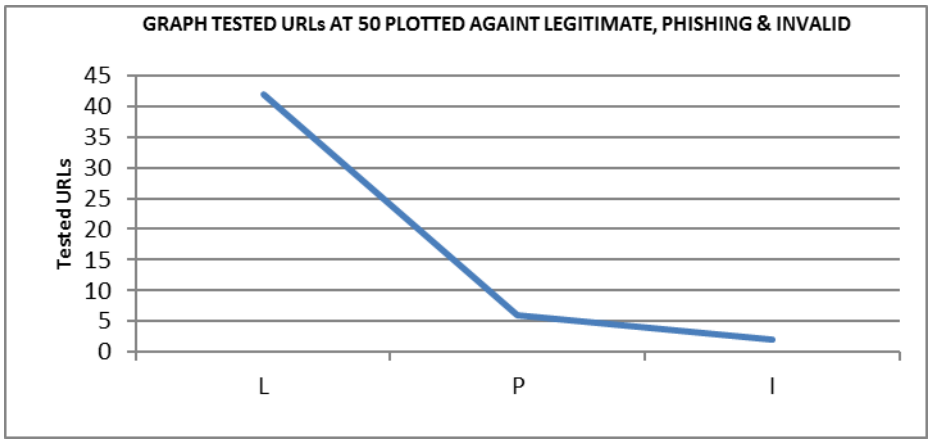

Figure 4: Graphical representation of Tested URL at 50 plotted against Legitimate, Phishing & Invalid URLs

When the number of tested URLs was increased, the application detects more phishing URLs than legitimate URLs based on the input. Figure 5 shows that when the total tested URLs was 128, the application detected 94 URLs correctly where 28 URLs was correctly classified as phishing and 6 URLs as invalid. The application could not classify it as legitimate or malicious since it is not registered as either.

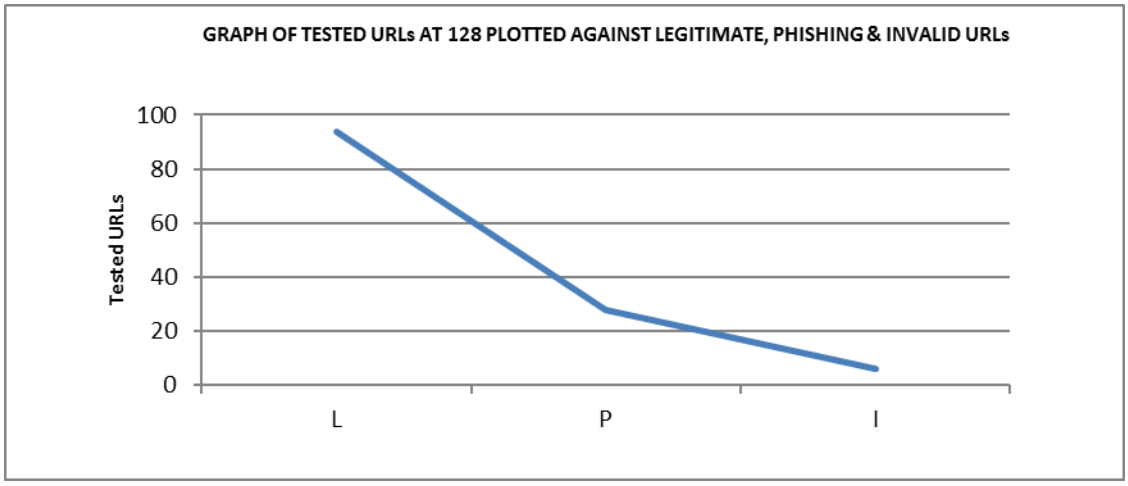

Figure 5: Graphical representation of URLs at 128 plotted against Legitimate, Phishing & Invalid

From Figure 6, as the total number of tested URLs increases to 300, the application correctly classified 109 URLs as legitimate, 171 classified as phishing and 20 are invalid.

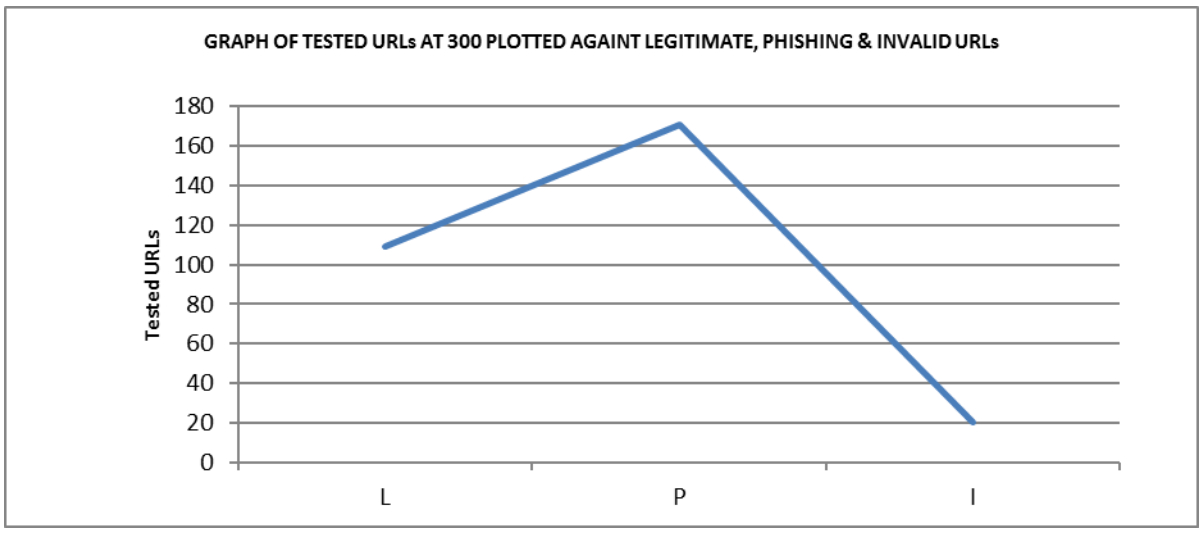

Figure 6: Graphical Representation of tested URLs at 300 plotted against Legitimate, Phishing & invalid URLs

The following analyses were carried out to test and know the effectiveness of the applications.

For Table 1, 112 URLs were classified as legitimate, but 110 URLs happened to be correctly classified while the remaining 2 URLs were not valid sites. 130 URLs are gathered as phishing which were correctly classified as such. The same method was applied for all the 5 rounds of testing, that is, for Tables 2 and 3.

Table 1. Experiment Result

	Legitimate URLs	Phishing URLs	Total
Z	112	130	242

t	110	130	240

Table 2. Experiment Result

	Legitimate URLs	Phishing URLs	Total
Z	205	237	442
t	202	237	439

Table 3. Experiment Result

	Legitimate URLs	Phishing URLs	Total
Z	255	301	556
t	253	301	554

Where "Z" is the total number of submitted URLs and "t" is the number of correctly classified URLs. Graph of Total Number of Submitted Legitimate URLs plotted Against Correctly Classified Legitimate URLs.

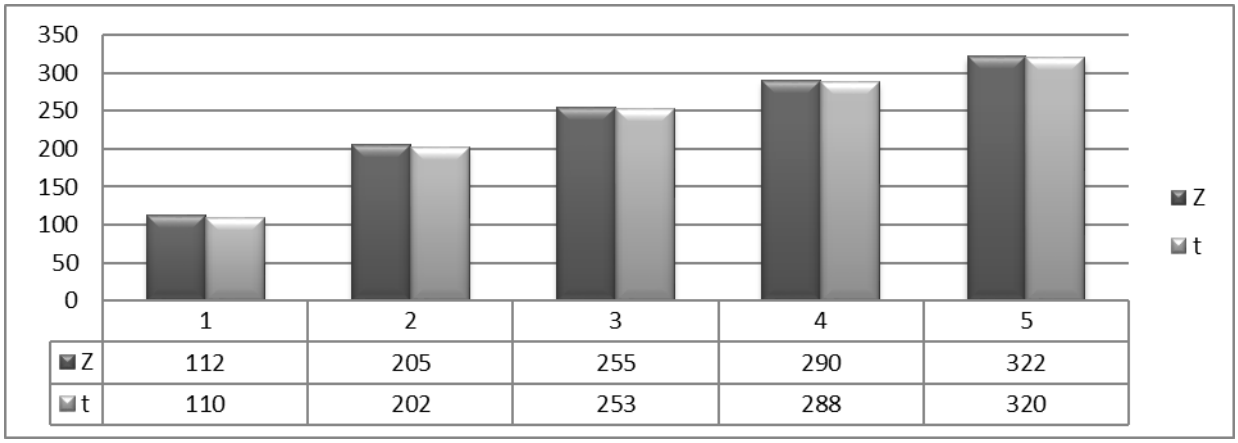

Figure 7: Graph of Total Number of Submitted URL Plotted Against Correctly Classified Legitimate URLs

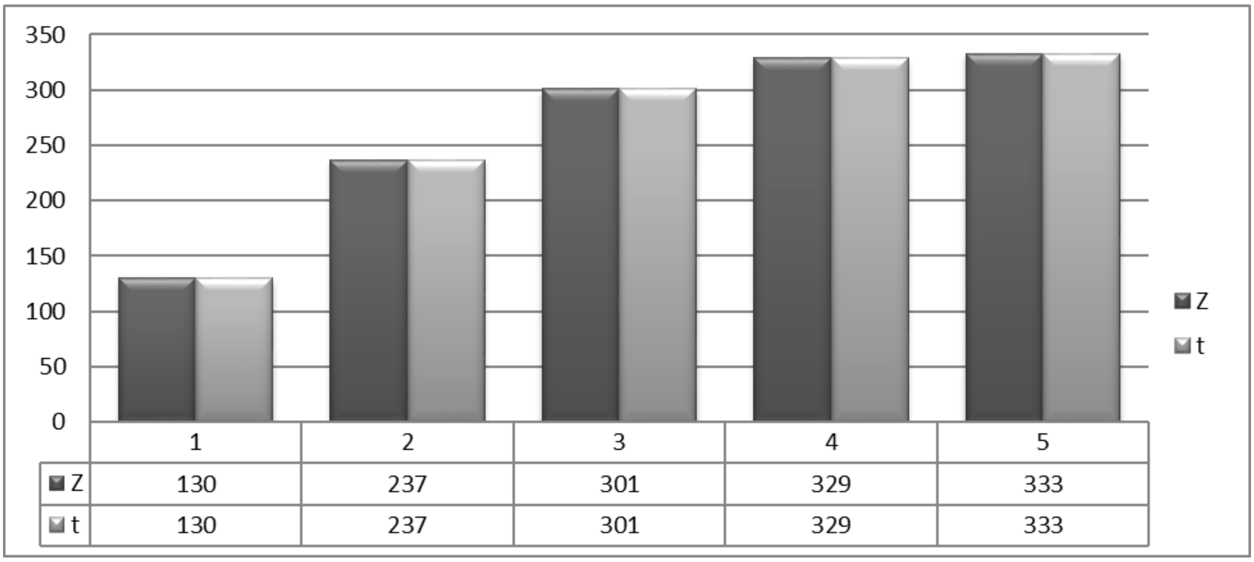

Figure 8: Graph of Total Number of Submitted URL Plotted Against Correctly Classified Phishing URLs

3.5 Evaluation Metrics

The following metrics were used in this work.

1. *True positive*: measures the rate of phishing sites (Ph) that are correctly classified as phishing. It is denoted as:

   TPR= $\frac{Ph \rightarrow Ph}{Ph + Fneg}$ … … … … … …. …………………………………………..(3)

Where F_{neg} = this is the number of phishing site mistakenly categorized as legitimate site

$$TPR = \frac{130}{130+0} \quad TPR = 1 * 100 = 100\%$$

2. *True Negative Rate (TNR):* this determines and measures the rate of legitimate sites (O) that is accurately categorized as legitimate sites. It is denoted as:

   TNR = $\frac{c \rightarrow o}{o + Fpos}$…………………………..…………………………………..…… (4)

Where F_{pos} = the number of legitimate site that is falsely categorized as phishing site

$$TNR = \frac{110}{110 + 2}, \quad TNR = \frac{110}{112}, \quad TNR = 0.98*100 = 98\%$$

3. *False Positive Rate:* Measures the rate of legitimate sites (O) wrongly categorised as phishing sites (P_h). it is denoted as:

   FPR = $\frac{o \rightarrow Ph}{Fpos+Tpos}$……………………………………………………………….. (5)

Where T_{pos} = the number of phishing site accurately categorised as phishing site

$$FPR = \frac{2}{2 + 130} = FPR = \frac{2}{132}, \quad FPR = 0.01515*100 = 1.5\%$$

4. *False Negative Rate:* measure the rate of phishing sites (Ph) falsely categorised as legitimate site (O). it is denoted as:

$$FNR = \frac{o \rightarrow Ph}{Fneg + Tneg} \quad \ldots \quad \ldots \quad \ldots \ldots \ldots \ldots \ldots \ldots (6)$$

Where T_{neg} = the number of legitimate sites accurately categorized as legitimate sites

$$FNR = \frac{0}{0+110}, \ FNR = 0 * 100\% = 0\%$$

5. *Accuracy:* is the measure of overall rate of classified sites in relation to the sum of the actual or correctly classified legitimate sites and phishing sites. It is denoted as:

$$Acc = \frac{(Ph \rightarrow O)}{Ph + Tpos + O + Tneg} \ldots \ldots \ldots \ldots (7)$$

$$Acc = \frac{130 + 110}{130 + 0 + 110 + 2}, \quad Acc = \frac{240}{242} = 0.991 * 100 = 99.1\%$$

4. Conclusion

Sequel to the damages done by phishing attacks in the cyber world to various legal and authorize activities of internet user, this project has therefore successfully carried out a method for detecting phishing sites with the aim of guiding, guarding and preventing internet users from falling prey of cybercriminals. In this solution, two main phases are involved: the pre-filtering phase and classification phase. For the pre-filtering phase, a special attention is given to the consistency between potential identities and second level domain names. The classification phase on other hand, breaks it further into many features vis-à-vis ratio of the found domain, position of the domain token and the randomness of the URL; all these have their attention focused on what is contained in the URL. Finally, content of the webpages and the content between the URL are examined by the conceptual similarity. If the URL of a webpage as gone through all these features and being considered phishing, the application further check PhishTank database to verify if the website is present then the site is a phishing site. URls that possess features similar to phishing and legitimate site are updated in the invalid database as they cannot be classified by the application. This approach is web-based that uses the content of the webpage and PhishTank database for classification. Previous knowledge of the websites is not necessary as the application is well programmed and easy to operate. As a future work, more features can be added that are not confined to the URLs in order to improve performance.

References

(1) Garera, S., Provos, N., Chew, M., and Rubin, A. D. (2007). A framework for detection and measurement of phishing attacks. Paper presented at the Proceedings of the 2007 ACM workshop on Recurring malcode. Pp 1-8

(2) Xiang, G., and Hong, J. I. (2009). A hybrid phish detection approach by identity discovery and keywords retrieval. International World Wide Web Conference Committee (IW3C2). Pp 1-10

(3) Abbasi, A., and Chen, H. (2009b). A comparison of tools for detecting fake websites. Computer,42(10), 78-86

(4) Afroz, S., and Greenstadt, R. (2009). Phishzoo: An automated web phishing detection approach based on profiling and fuzzy matching: Technical Report DU-CS-09-03, Drexel University.
(5) Zhang, J., Ou, Y., Li, D., and Xin, Y. (2012). A Prior-based Transfer Learning Method for the Phishing Detection. Journal of Networks, 7(8), 1201-1207
(6) Ma, J., Saul, L. K., Savage, S., and Voelker, G. M. (2009). Identifying suspicious URLs: an application of large-scale online learning. n Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada, 2009. Pp 1-8
(7) Shreeram, V., Suban, M., Shanthi, P., and Manjula, K. (2010). Anti-phishing detection of phishing attacks using genetic algorithm. 2010 IEEE International Conference on Communication Control and Computing Technologies (ICCCCT), Ramanathapuram, India , 2010 IEEE International Conference on. Pp 447-450
(8) Afroz, S., and Greenstadt, R. (2011). Phishzoo: Detecting phishing websites by looking at them. 2011 Fifth IEEE International Conference on Semantic Computing (ICSC), 2011, Palo Alto, CA, USA, pp 368-375
(9) Alnajim, A., and Munro, M. (2009). An Approach to the Implementation of the AntiPhishing Tool for Phishing Websites Detection. International Conference on Intelligent Networking and Collaborative Systems, 2009. INCOS '09. Barcelona, Spain. Pp 105-112.
(10) Basnet, R., Mukkamala, S., and Sung, A. (2008). Detection of phishing attacks: A machine learning approach. Soft Computing Applications in Industry, 373-383
(11) Chen, J., and Guo, C. (2006). Online detection and prevention of phishing attacks. Paper presented at the First International Conference on Communications and Networking in China, 2006 (ChinaCom'06), pp 1-7.
(12) Chou, N., Ledesma, R., Teraguchi, Y., Boneh, D., and Mitchell, J. C. (2004). Clientside defense against web-based identity theft. 11th Annual Network and Distributed System Security Symposium (NDSS'04). Pp 1-16.
(13) Dunlop, M., Groat, S., and Shelly, D. (2010). GoldPhish: Using Images for ContentBased Phishing Analysis. The Fifth International Conference on Internet Monitoring and Protection (ICIMP), pp 123-128
(14) Fu, A. Y., Wenyin, L., and Deng, X. (2006). Detecting phishing web pages with visual similarity assessment based on earth mover's distance (EMD). Dependable and Secure Computing, IEEE Transactions on, 3(4), 301-311
(15) Gabber, E., Gibbons, P. B., Kristol, D. M., Matias, Y., and Mayer, A. (1999). Consistent, yet anonymous, Web access with LPWA. Communications of the ACM, 42(2), 42-47.
(16) Liu, G., Qiu, B., and Wenyin, L. (2010). Automatic Detection of Phishing Target from Phishing Webpage. Paper presented at the Pattern Recognition (ICPR), 2010 International Conference on Pattern Recognition. Pp 4153-4156
(17) Liu, W., Deng, X., Huang, G., and Fu, A. Y. (2006). An antiphishing strategy based on visual similarity assessment. Internet Computing, IEEE, 10(2), 58-65.
(18) Martin, A., Anutthamaa, N., Sathyavathy, M., Francois, M. M. S., and Venkatesan, D. V. P. (2011). A Framework for Predicting Phishing Websites Using Neural Networks. arXiv preprint arXiv:1109.1074.
(19) Prakash, P., Kumar, M., Kompella, R. & Gupta, M., (2010). Phishnet: predictive blacklisting to detect phishing attacks. IEEE, in INFOCOM, 2010 Proceedings IEEE (pp. 1-5), pp. Pages 346-350.
(20) Roa, R. & Ali, S., (2015). PhishShield: A Desktop Application to Detect Phishing Webpages through Heuristic Approach.. S.l., Procedia Computer Science, 54, pp.147-156.
(21) Schneider, F., Provos, N., Moll, R., Chew, M., and Rakowski, B. (2007). Phishing Protection Design Documentation. URL: https://wiki.mozilla.org/Phishing_Protection:_Design_Documentation. Date accessed: 20-08-2017.
(22) Suen, C. Y., Nadal, C., Legault, R., Mai, T. A., and Lam, L. (1992). Computer recognition of unconstrained handwritten numerals. Proceedings of the IEEE, 80(7), 1162-1180
(23) Yue, Z., Jason, H. & Cranor, L., (2007). CANTINA: A Content-Based Approach to Detection of Phishing Websites, Banff, Alberta, Canada. Volume ACM 978-1-59593-654-7/07/0005.
(24) Zhang, Y., Egelman, S., Cranor, L., and Hong, J. (2006). Finding phish: Evaluating anti-phishing tools. Conference Proceeding at Carnegie Mellon University, pp 1-16.
(25) Zhang, Y., Hong, J. I., and Cranor, L. F. (2007). Cantina: a content-based approach to detecting phishing web sites. The International World Wide Web Conference Committee (IW3C2). WWW 2007, May 8–12, 2007, Banff, Alberta, Canada. Pp 1-10
(26) Zhuang, W., Jiang, Q., and Xiong, T. (2012). An Intelligent Anti-phishing Strategy Model for Phishing Website Detection. 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW), 2012 , Macau, China. pp 51-56
(27) Chen, Y. Liu, H. Yu, Y. and Wang, P (2014) "Detect Phishing by Checking Content Consistency" 2014 IEEE 15th International Conference on Information Reuse and Integration (IRI), IEEE IRI 2014, August 13-15, 2014, San Francisco, California, USA. pp 109-116

(28) Azeez, N. A., and Iliyas, H. D. (2016). Implementation of a 4-tier cloud-based architecture for collaborative health care delivery. Nigerian Journal of Technological Development, 13 (1), 17-25.
(29) Azeez, N. A., and Venter, I. M. (2013). Towards ensuring scalability, interoperability and efficient access control in a multi-domain grid-based environment. SAIEE Africa Research Journal, 104 (2), 54-68.
(30) Azeez, N. A., Iyamu, T., and Venter, I. M. (2011). Grid security loopholes with proposed countermeasures. In E. Gelenbe, R. Lent, and G. Sakellari (Ed.), 26th International Symposium on Computer and Information Sciences (pp. 411-418). London: Springer.
(31) Azeez, N.A., and Lasisi, A. A. (2016). Empirical and Statistical Evaluation of the Effectiveness of Four Lossless Data Compression Algorithms. Nigerian Journal of Technological Development, Vol. 13, NO. 2, December 2016, 64-73.
(32) Nureni, A. A., and Irwin, B. (2010). Cyber security: Challenges and the way forward. Computer Science & Telecommunications, 29, 56-69.
(33) Azeez, N. A. (2012). Towards Ensuring Scalability, Interoperability and Efficient Access Control In a Triple-Domain Grid-Based Environment. Cape Town: University of the Western Cape.
(34) Ayofe, A.N, Adebayo, S.B, Ajetola, A.R, Abdulwahab, A.F (2010) "A framework for computer aided investigation of ATM fraud in Nigeria" International Journal of Soft Computing, Vol. 5, Issue 3 pp. 78-82
(35) Azeez, N.A, Olayinka, A.F, Fasina, E.P, Venter, I.M. (2015) "Evaluation of a flexible column-based access control security model for medical-based information" Journal of Computer Science and Its Application. Vol. 22, Issue 1, Pages 14-25
(36) Azeez, N. A., and Ademolu, O. (2016). CyberProtector: Identifying Compromised URLs in Electronic Mails with Bayesian Classification. 2016 International Conference Computational Science and Computational Intelligence (CSCI) (pp. 959-965). Las Vegas, NV, USA: IEEE.
(37) Azeez, N. A., and Babatope, A. B. (2016). AANtID: an alternative approach to network intrusion detection. The Journal of Computer Science and its Applications. An International Journal of the Nigeria Computer Society, 129-143.