IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 9, September 2020

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 9, September 2020 ||

Detection of Phishing Website Using Deep Learning Approach

Shivani Pramod Wagh[1], Himani Vilas Mahajan[1], Nikita Vikas Bhavsar [1], D.A.Meshram[2]

B. E Students, Department of Information Technology, RMD Sinhgad School of Engineering Warje, Pune, India [1]

Assistant Professor, Department of Information Technology, RMD Sinhgad School of Engineering Warje, Pune, India [2]

ABSTRACT: In today's world, the internet has brought tremendous change in e-commerce aspect of people's live. However it is prone to the wide variety of security attacks. One of the most vulnerable security threat which poses a problem to the internet users is phishing.

Phishing is an attack made to steal the sensitive information of the users such as passwords, PIN, social security number etc. In phishing attack, the attacker creates a fake website to make the users click it and steal the sensitive information of users. There are many ways to detect the phishing website or phishing URLs. Phishing website contains many cues both in their content part as well as in browser based security indicators. There are many solutions provided by several researchers in which no single approach evolves as an effective method to detect phishing attacks.

KEYWORDS: Phishing sites, URL, Mail spammers, DCDA (Dynamic Category Decision Algorithm), RNN, Question-Answer Security.

I. INTRODUCTION

There are number of users who purchase products online and make payment through e- banking. There are e- banking websites who ask user to provide sensitive data such as username, password or credit card details etc. often for malicious reasons. This type of e-banking websites is known as phishing website. In order to detect and predict e-banking phishing website, we proposed an intelligent, flexible and effective system that is based on using classification Data mining algorithm. We implemented classification algorithm and techniques to extract the phishing data sets criteria to classify their legitimacy.

The e-banking phishing website can be detected based on some important characteristics like URL and Domain Identity, and security and encryption criteria in the final phishing detection rate. Once user makes transaction through online when he makes payment through e-banking website our system will use data mining algorithm to detect whether the e-banking website is phishing website or not. This application can be used by many E- commerce enterprises in order to make the whole transaction process secure. Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms. With the help of this system user can also purchase products online without any hesitation A bit of the cases are gathering sees, capable news or meeting assertions. Phishing itself is another idea, yet it's undeniably utilized by the assailants for example the phishers to take your own data and perform business and social wrongdoings as of late. Inside four to five years the quantities of phishing assaults have expanded significantly. Phishing assaults are usually utilized and are anything but difficult to execute on their objective. Ordinarily phishing assault abuses the social designing to draw the unfortunate casualty through sending a caricature interface by diverting the injured individual to a phony website page. [5] The caricature connect is put on the famous pages or sent by means of email to the person in question. The phony site page is made like the authentic website page. In this manner, instead of guiding the injured individual solicitation to the genuine web server, it will be coordinated to the aggressor server.[7]

- Impacts of Phishing:

1. Denial of Service Implications. Phishing: Denial of Service (DoS)Implications.
2. Financial Losses. Financial Losses from Phishing.
3. Reputational Damages. Reputational Damages

- Types of phishing:

1. Vishing. Vishing refers to phishing done over phone calls.
2. Smishing. SMS phishing or SMiShing isone of the easiest types of phishing attacks.
3. Search Engine Phishing.
4. Spear Phishing.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 9, September 2020 ||

5. Whaling.

II. MOTIVATION

- The phishing issue is wide and no single silver- slug arrangement exists to moderate every one of the vulnerabilities viably, in this manner different procedures are regularly actualized to alleviate explicit assaults.
- User can make online payment securely.
- Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms.
- With the help of this system user can also purchase products online without any hesitation
- The point of the proposed framework is recognizing the most extreme number of phishing assaults utilizing Artificial Neural Network calculation.

III. SYSTEM OVERVIEW

At the point when Costumer first time login to System that time all nuances IP address, current device, OS, time, region set aside customer log records at bank server for future peculiarity recognizable proof. Next time when customer need to login that time these nuances differentiated and customers currentnuances at whatever point facilitated by then get to permitted regardless security Questions asked concerning whether okay with this then n at precisely that point get to surrendered. Client login to the system, after login customer gets affirmation agree to access or view structure. System is liable for offering access to customer by basically organizing quirks at whatever point composed then okay regardless check for character. We have to give perfect approach to manage electronic setting aside cash with the help of anomaly-based area and expectation of phishing ambushes

SYSTEM ARCHITECTURE

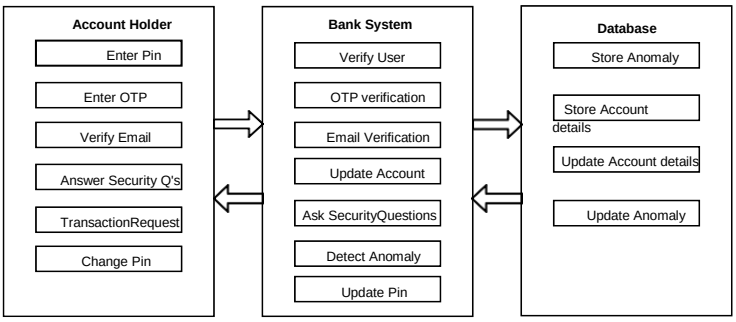

Fig. 01 System architecture

IV. ALGORITHMS

Phishing Detection:

Phishing URL Detection can be done by proactive or by reactive means. On the reactive end, we find services such as Google Safe Browsing API. This type of services expose a blacklist of malicious URLs to the queried.. Blacklists are constructed by using different techniques, including manual reporting, honeypot sorby crawling the web in search of known phishing characteristics. For Example, browser make use of blacklists to block access upon reaching the URLs contained in them. One drawback of such reactive method is that in order for phishing URL to be blocked, it must be previously included in the blacklist.

Proactive methods mitigate this problem by analyzing the characteristics of web page in real time in order to assess the potential risk of web page. Risk assessment is
done through classification model. Some of the machine learning methods that have been used to detect phishing include: support vector machine, streaming analytics, gradient boosting, random forests, online incremental learning

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 9, September 2020 ||

and neural network. The application of machine learning techniques for URL classification has been gaining attention.

Uniform Resource Locator Structure:

The URL is string representation for a resource available on the internet. A URL is written as follows:

<scheme>:<scheme-specific-part>

The scheme specifies the resources access mechanism or network location (i.e.http,ftp,mailto), while the rest of the URL may vary depending on the scheme selected.

1. RNN Algorithm:

In RNN, instead of manually extracting the features, we directly learn a representation from the URLs character sequence. Each character sequence exhibits correlations, that is, nearby characters in a URL likely to be related to each other. These sequential patterns are important because they can be exploited to improve the performance of the predictors..

RNN (Recurrent Neural Network) is primarily used in NLP. For instance, an application of RNN is in language modelling or text generation. These kind of tasks demands you to understand the semantics and syntactic form of the sentences. This is because, the RNNs cannot take care of the long-term dependency of a, sentence. A neural network is a bio-inspired, machine- learning model that consists of a set of artificial neurons with connections between them. Recurrent Neural Networks (RNN) are a type of neural network able to model sequential patterns. The distinctive characteristic of RNNs is that they introduce the notion of time to the model. This allows them to process sequential data one element at a time and to learn their sequential dependencies of RNNs is that they introduce the notion of time to the model. This allows them to process sequential data one element at a time and to learn their sequential dependencies

1. Initialize loads arbitrarily

2. Give the model a roast pair (input scorch and target burn. The objective roast is the scorch the system should figure, its the following singe in our grouping)

3. Forward pass (We compute the likelihood for each conceivable next roast as per the condition of the model, utilizing the paramters)

4. Measure mistake (the separation between the past likelihood and the objective scorch)

5. We ascertain angles for every one of our parameters to see their effect they have on the misfortune (backpropagation through time)

6. update all parameters toward the path by means of inclinations that help to limit the misfortune

1. Repeat! Until our misfortune is little AF

Given a sequence of words $\omega_{1:n} = \omega_1, \ldots, \omega_n$, where each is associated with an embedding vector of dimension d. A 1D convolution of width-k is the result of moving a sliding- window of size k over the sentence, and applying the same convolution filter or kernel to each window in the sequence, i.e., a dot- product between the concatenation of the embedding vectors in a given window and a weight vector u, which is then often followed

1. Considering a window of words

$\omega_i, \omega_{i+1}, \ldots, \omega_{i+k}$the concatenated vector of the ith window is then:

2. ……. (1)

3. The **convolution filter** is applied to each window, resulting in scalar values r_i, each for the ith window:

4. …….. (2)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 9, September 2020 ||

5. In practice one typically applies more filters,
$u_1, \ldots, u_l$, which can then be represented as a vector multiplied by a matrix U and with an addition of a bias term b:

$$r = g(x \cdot U + b) \quad \ldots\ldots (3)$$

4. non-linear activation function $g.r = g(x \quad U + b)$ ….. (3)

5.

7. with$r_i \in R^l$, $x_i \in R^{k \times d}$, $U \in$

$R^{k \times d \times l}$ and $b \in R^l$

A neural network is bio-inspired machine learning model that consists of set of artificial neurons with connections between them. Recurrent Neural Network are a type of neural network that is able to model sequential patterns. The distinctive characteristics of RNNs is that they introduce the notion of time to the model., which it in turn allow them to process sequential data one element at a time and learn their sequential dependencies.

2.**DCDA (Dynamic Category Decision Algorithm):**

Dynamic decision-making (DDM) is interdependent decision-making that takes place in an environment that changes over time either due to the previous actions of the decision maker or due to events that are outside of the control of the decision maker. Dynamics is defined as the branch of mechanics that deals with the effect of outside forces on something. A genetic algorithm is used to identify the most important feature subset for prediction. Principal component analysis is used to remove irrelevant and redundant features. In multidimensional learning tasks, where there are multiple target variables, it is not clear how feature selection should be performed

1: p0 = CNN-LSTM(ui)

2: p1 = 1-p0

3: maxi = max (p0, p1) 4: mini = min (p0, p1)
5: if (maxi / mini ||! (@)) imaxi mini access (ui) then
6: if 10 (p1>p) then

7: return " phishing 8: else return " legitimate 9: end if
10: else return multidimensional_features (ui)
11: end if

V. RESULT AND DISCUSSION

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3- 2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server. To build up an application for prevention from Attackers used phishing technique to steal personal confidential information and used different techniques to detect these types of attack using feature engineering which requires previous knowledge of attacks and requires large amount of time

	Algorithm	Accuracy
Existing Algorithm	SVM	69%
Proposed Algorithm	DCDA	93%

Fig. 2 Accuracy associated with algorithm detection

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 9, September 2020 ||

VI. CONCLUSION

The examination the final answer for tackle issue appropriately and the inquiry "How to recognize and forestall conceivable unapproved login endeavors through taken subtleties from a phishing assault in an internet banking framework? "was finally replied by the proposed arrangement utilizing three components effectively.
The three instruments can be classified an oddity-based location, IP address identification and gadget identification. The general framework won't just identify the unapproved login endeavors yet additionally forestall it, notified to approved clients and shield web based financial clients from fraudsters.

REFERENCES

1. Jian Mao, Wenqian Tian, Pei Li, Tao Wei, Zhenkai Liang, “Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity”, IEEE Access Volume:5
2. Zou Futai, Gang Yuxiang, Pei Bei, Pan Li, Li Linsen, “Web Phishing Detection Based on Graph Mining”, 2016 2nd IEEE International Conference on Computer and Communications(ICCC)
3. Nick Williams, Shujun Li, “Simulating human detection of phishing websites: An investigation intotheapplicabilityofACT-Rcognitivebehaviour architecture model”,
4. Xin Mei Choo, Kang LengChiew,DayangHananiAbangIbrahim,Nadiana tra Musa, San Nah Sze, Wei King Tiong, “Feature- Based Phishing Detection Technique”, 2016 3rd International Conference on Computing for Sustainable Global Development(INDIACom)
5. Giovanni Armano, Samuel Marchal and N. Asokan, “Real-Time Client-Side Phishing Prevention Add-on”, 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)
6. Trupti A. Kumbhare and Prof. Santosh V. Chobe, “An Overview of Association Rule Mining Algorithms”,
7. S.Neelamegam, Dr.E.Ramaraj, “Classification algorithm in Data mining: An Overview”, International Journal of P2P Network Trends and Technology (IJPTT) - Volume 3 Issue 5 September to October2013
8. Varsharani Ramdas Hawanna, V. Y. Kulkarni and R. A. Rane “A Novel Algorithm to Detect Phishing URLs.”, 2016 International Conference on Automatic Control and Dynamic Optimization Techniques(ICACDOT)
9. Jun Hu, XiangzhuZhang,Yuchun Ji, Hanbing Yan,LiDing,JiaLiandHuimingMeng“Detecting Phishing Websites Based on the Study of the Financial Industry Webserver Logs.” , 2016 3rd International Conference on Information Science and Control Engineering(ICISCE)
10. Zheng Dong, Apu Kapadia, Jim Blythe†and L. Jean Camp,“Beyond the Lock Icon: Real-time Detection of Phishing Websites Using Public Key Certificate”, IEEE 2015.
11. Samuel Marchal, N. Asokan,“On Designing and Evaluating Phishing Webpage Detection Techniques forthe Real World”, USENIXWorkshop on Cyber SecurityExperimentation and Test - Baltimore, United States13 Aug2018.
12. Saad Tayyab, Asad Masood, “A Review: Phishing Detection using URLs and Hyperlinks Information by Machine Learning Approach”, IJCSMC, Vol. 8, Issue. 3, March 2019, pg.345–351
13. Robat Das, Md. Mukhter Hossain, Shariful Islam, AbujarrSiddiki,“ Learning a Deep Neural Network for Predicting Phishing Website”, Spring April25,2019.
14. Shireen Riaty, Ahmad Sharieh, Hamed Al Bdour, “Enhance Detecting Phishing Websites Based on Machine Learning Techniques of Fuzzy Logic with Associative Rules ”,
15. Bo Wei, Rebeen Ali Hamad, Longzhi Yang, Xuan He , Hao Wang , Bin Gao and Wai LokWoo
16. , “A Deep-Learning-Driven Light-Weight Phishing Detection Sensor” , Sensors 2019
17. DI XUE, JINGMEI LI, TU LV, WEIFEIWU,ANDJIAXIANGWANG,“MalwareClassification Using Probability Scoring and MachineLearning”,IEEE Access VOLUME 7, 2019
18. Arun Kulkarni1, Leonard L. Brown, “Phishing Websites Detection using Machine Learning”, International Journal of Advanced Computer Science and Applications, Vol. 10, No. 7,2019
19. Ping Yi , Yuxiang Guan, Futai Zou, Yao Yao, Wei Wang, and Ting Zhu, “Web Phishing Detection Using a Deep Learning Framework”, Wireless Communications and Mobile Computing Volume 2018,
20. R. Kiruthiga, D. Akila, “Phishing Websites Detection Using Machine Learning”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S11,
21. September 2019
22. PurviPujara, M. B.Chaudhari, “Phishing Website Detection using Machine Learning : A Review”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT | Volume 3 | Issue 7 | ISSN :2456-3307
23. Ms. Sophiya Shikalgar Dr. S. D. Sawarkar , Mrs. Swati Narwane ,“Detection of URL based Phishing Attacks using Machine Learning”, International Journal of Engineering Research& Technology(IJERT)Vol.8Issue11,November-2019
24. RamBasnet, Srinivas Mukkamala, and Andrew H. Sung,“Detection of Phishing Attacks: A Machine Learning

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 9, Issue 9, September 2020 ||

Approach", Springer-Verlag Berlin Heidelberg 2008
25. DOYEN SAHOO,CHENGHAO LIU, STEVENC.H. HOI, "Malicious URL Detection using Machine Learning: A Survey",Vol. 1, No.1, Article.Publicationdate:August2019
26. Eint Sandi Aung, ChawThet Zan and HayatoYAMANA, "A Survey of URL-based Phishing Detection", DEIM Forum2019
27. Brad Wardman, Tommy Stallings, Gary Warner, Anthony Skjellum ,"High-Performance Content-Based Phishing Attack Detection"
28. Neda Abdelhamid,FadiThabtah,Hussein Abdel-jaber,"Phishing Detection: A Recent Intelligent Machine Learning Comparison basedon Models Content and Features",©2017IEEE
29. Amol D Wakhare1,V SKarvande ,"DETECTION OF PHISHING WEB SITES BASED ON FEATURE CLASSIFICATION AND EXTREMELEARNINGMACHINE",|Volume4|| Issue 10 || OCTOBER 2019 || ISO3297:2007 Certified

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details