Juni Khyat
(UGC Care Group I Listed Journal)
ISSN: 2278-4632
Vol-10 Issue-6 No. 3 June 2020

A Methodology to predicate the phishing vulnerabilities using Decision Tree

N Md Jubair Basha [1], D Ramadevi [2], Gollapudi Sai Jayasri [3], D Mounika [4]
CSE Departmen, Kallam Haranadhareddy Institute of Technology, Guntur, A.P.
[1] nawabjubair@gmail.com
[2] ramadevidevireddy9@gmail.com
[3] jayashree.gollapudi@gmail.com
[4] devarasettymounika2@gmail.com

Abstract— **Web application attacks are increasing and exploiting the security of users. The related work includes the problem of detecting the attacks can be prevented using blacklist and fuzzy role techniques. But these techniques lacks with dealing of phishing attacks. The problems in the existing literature can be prevented using decision tree algorithms. But Decision Tree algorithm is used efficiently to detect the attacks and secure the user. This paper presents a methodology uses an open source Web Application Firewall- Mod Security which is an internet technology helps preventing the attacks. The proposed approach of WAF with the algorithms of machine learning to efficiently detect the attacks and secure the user. Machine learning methods learns the attack from previous attacks according to the previous results and block or bypass the Web Applications Firewall. This paper focuses on SQL Injections and Phishing vulnerabilities and prevent attackers to easily deceive them.**

Keywords*— *phishing attack, decision tree, vulnerability, firewall

I. INTRODUCTION

Phishing is a type of cyber security attack often used to steal user data, including login credentials by sending spam email or any other channel. Phishing is one of the most popular technique because it is easy for the attacker to trick a user by sending any spam or malicious URL which seems similar to a known website. The malicious links are developed in such a way that they represent a true organisation with its fake logo and other true contents. The users click the malicious link and are exploited by the attackers easily.

This paper focus on these attacks which can be harmful for the users or can bypass any firewall. SQL Injections attacks target the database server using malicious code and gain critical data stored in the database. These attacks are done by exploiting SQL vulnerabilities [12,14] allowing the SQL server to run the harmful code.

This paper presents the related work in Section II, proposed methodology in Section III, proposed architecture in Section IV, Decision Tree algorithm to predict phishing vulnerabilities is presented in Section V, respective results in Section V and Concludes the paper with respective references.

II. RELATED WORK

A Web Application is an application software that uses online web browser to perform different task over the internet [1,2]. Millions of businesses and common people use the internet for different purpose like cost-effective communication channels, storing data, making transactions, accessing social networking sites. These Web Applications are easily exploited and attacked by hackers and crackers [3,4,5]. However, these attacks can be prevented by using the concepts of Web Application Firewall [10] and Machine Learning. WAF enables the user to access the real time Web Application monitoring and its access control.

Firewall is a network security system that can control and monitor the traffic on the network based on some rules that are predefined for the security purpose. In actual the firewall establishes a barrier between a trusted network and an untrusted network which can be an internet. There is category of firewall- Network firewall and host-based firewall. Network based firewall works on LANs, WANs and intranets. These can be a software application running on a hardware or a hardware firewall itself. Firewall can also possess others functions such as VPN servers. Host based firewall are positioned on the network node for controlling network traffic to and from the machine. These firewalls are divided into types which are network layers or

DOI : 10.46528/JK.2020.V10I06N03.027

ISSN: 2278-4632
Vol-10 Issue-6 No. 3 June 2020

packet filters, Applications layers, Proxys and network address translations. Packet filters are positioned on low level of TCP/IP Protocols, basically disapprove the packet to pass through firewall unless they are matched with predefined rule set.

A Proxy server also act as firewall by responding to packets at the receiving end in the form of an applications while blocking others packets. Network address translations is a function of firewall and protect the host that have addresses in the private address range. The last type of firewall which is Applications layer firewall is a type of firewall that defines Web Applications firewall which work in the application level of TCP/IP stack that obstruct all packet travelling in and out of an application. Mod Security is an open source tool kit for real time monitoring of web application and its access control. The four guiding principles on which these tool kits are based are flexibility, passiveness, predictability and quality over quantity. It is cross platform WAF module which enables web application defenders to produce the visibility into the traffic of HTTP and provides rules language and API for implementing other security protection. As discussed, these Web Application Firewall bypass the cyber security attacks that may try to cause harm.

Most common types of attacks that are seen now-a-days are- Malware, XSS, DoS, Session hijacking, Information Reuse, Phishing and SQL Injections [11,13,15]. Malware are harmful software such as ransom ware and viruses. Malware can harm the computers and take control of the machine to silently monitor the action and keys strokes that can release the confidential data from the user. Cross-Site Scripting can lead to attack and target a website user with a loop hole and targets its data such as credentials and financial data by injecting harmful code into a website.

In Denial of Service a flood is created on a website to increase the traffic more than it was built to handle and make the website content unavailable to the users accessing it. Session Hijacking will be seizing of session through by catching the session id and propose as a PC making a demand which enable them to login and obtain entrance as an unapproved client. Most people reuse the same credentials which has been used previously. The Credential Reuse is the easiest type of attack in which the attackers have a collection of password and username from a breached website which is easily available on black market website. These credentials can be the same that a user is using currently and the attacker can gain the access to his/her e-mail, bank account, social networking, etc.

The Phishing attack pretends to a trustworthy website or an email where the users are asked to enter some personal information's or credentials. In these attacks the attacker mimics a trustworthy website which has a legitimate looking and make a trap to capture the details.

The phishing attacks in this modern internet world has spread rapidly. These phishing attacks can lead to many financial and similar losses. It is very difficult to trace the hacker. Thus, the first solution for preventing these types of attacks start from the awareness from the user which may not be a successful method. The non-technical method which is a legal solution to the problem. In many countries which requires the task to trace the hackers which is not 100% successful approach.

Two approaches are used in technical method-

- ***Blacklist technique*** is a database of pre-established phishing techniques or websites. Thus, it doesn't deal with each and every phishing website.
- ***Fuzzy role*** approach which includes gathering of features. These techniques have not seemed to be effective in preventing and detecting the phishing attack

The problem of detecting the attacks can be prevented using blacklist and fuzzy role techniques. But these techniques lack with dealing of phishing attacks. It has been analysed that phishing attacks can be prevented using one of the techniques of machine learning [6,7,8,9] decision tree algorithm effectively.

III. Proposed Methodology

This paper presents the concept of machine learning to train the system and detect these attacks efficiently and securely with following steps.

Step 1: The spam emails or phishing emails which can be detected using decision tree algorithm.

Step 2: Decision tree is used for detecting phishing attacks since it is considered as improved version of nested if-else where each feature is checked one by one.

DOI : 10.46528/JK.2020.V10I06N03.027

ISSN: 2278-4632
Vol-10 Issue-6 No. 3 June 2020

Step 3: These machine learning techniques can be grouped into the following features, those are URL-based, Domain based, page based and content based.
Step 4: The detection is a classification problem. Therefore, the data sets for phishing is done from an open source, phish tank which is a commonly used data source in academic studies.
Step 5: Apply the algorithm used in detecting these attacks in decision tree algorithm, which is a simple and powerful.

IV. Proposed Architecture

The proposed methodology needs the respective architecture which includes the following process. This following process executes to avoid and prevent the phishing attacks occurring in sequence. The architecture includes the following steps as follows:

1. Attacker will attack the web pages by passing the various credentials of different users.
2. After passing the credentials, Firewall will provide the ModSecurity by setting various String Patterns from the Training data.
3. A Machine Learning approach is used here to detect path avoid the permissions.
4. Later on, the same data is replicated to the various application, internal and database servers.

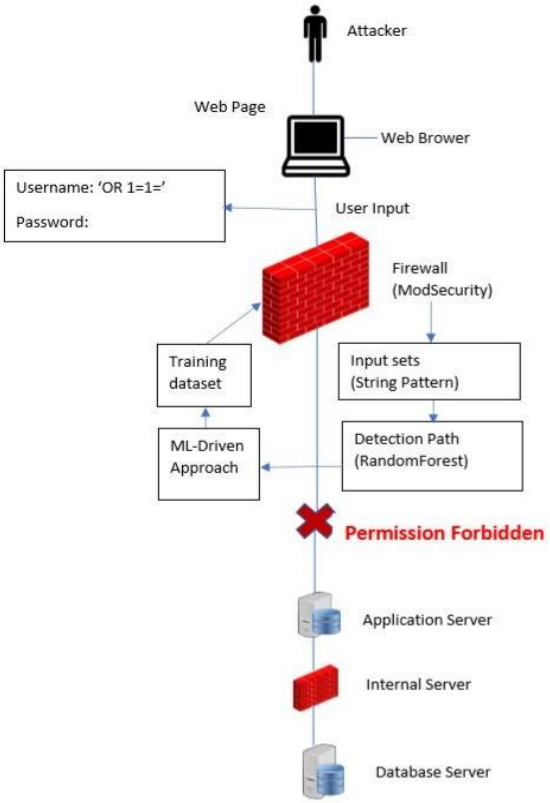

Fig. 1. Prevention of phishing attack

V. Decision Tree Algorithm to predict Phishing Vulnerabilities

A decision tree is a graph that uses a branching method to illustrate every possible outcome of a decision. Decision tree is commonly used tool in data mining for deriving a strategy to reach a particular goal, it's also widely used in machine learning. Decision tree is advantage compared with other approaches of being meaningful and easy to interpret. The Machine Learning algorithm has its own working mechanism.

The algorithm used is decision tree algorithm which is simple and powerful used to predict phishing vulnerabilities. The first task is the need of labelled instances to create detection mechanism. This is done by done by two classes- Phishing and legitimate. The decision tree can be considered as improved version of nested if-else where each feature is checked one by one. Generating a tree is the super structure of the

DOI : 10.46528/JK.2020.V10I06N03.027

ISSN: 2278-4632
Vol-10 Issue-6 No. 3 June 2020

mechanism. The length of the tree is checked when an example arrives and other features are checked as per result.

Decision tree algorithm

```
Tree -Learning(TR,Target,Attr)
    TR:training examples
    Target: target attribute
    Attr:set of descriptive attributes
{
    Create a root node for the tree.
   If TR have the same target attribute value ti.
         Then return the single node tree i.e Root with
          Target attribute=ti
      If Attr=empty
         Then return the single node tree i.e; Root
            With most value of target in TR
      Otherwise
      {
     Select attribute A from Attr that  best
     Classify TR based on an entropy-based measure
        Set A the attribute for root
         For each legal value of A, vi,do
         {
            Add a branch below Root, corresponding
             to A=Vi
            Let TRvi be the subset of TR that have
              A=Vi
           If TRvi be empty.
              Then add a leaf node below the branch
              with target value =most common value
             of Target in TR
             Else below the branch, add the
             subtree learned by
                  Tree Learning(TRvi,Targwt,Attr-{A})
         }
}
Return(Root)
}
```

VI. RESULTS

As per the result of ROC (Receiver OperativeCharacteristics – a graphical comparison plot between sensitivity and specificity) curve which is used to plot the result of TP (True Positive) vs FP (False Positive) which identify all the positive examples and is a perfect classifier to classify positive cases and negative cases efficiently. The accuracy of the test depends on the classification of the group and is measured by the area under theROCcurve.

Figure 2 presents the respective analysis with the phishing attacks by predicting the phishing vulnerabilities using decision tree algorithm. The accuracy is measured and presented. Figure 3 shows the respective accuracy measure in concern with the NLP Feature and VectorWorld.The comparison has carried out and predicted the respective phishing vulnerabilities in the web application.

DOI : 10.46528/JK.2020.V10I06N03.027

ISSN: 2278-4632
Vol-10 Issue-6 No. 3 June 2020

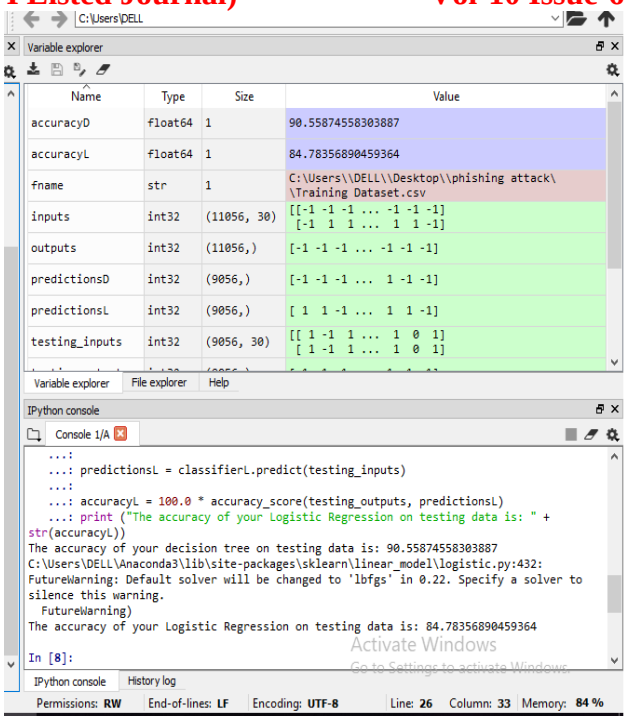

Fig. 2. URL based Phishing Evaluation

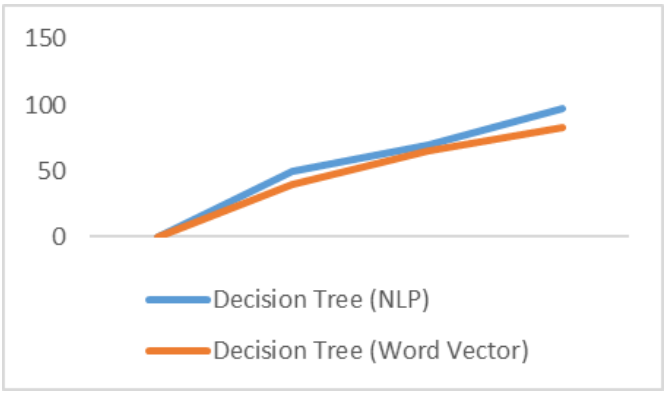

Fig. 3. Comparison of NLP Feature and World Vector

According to above graph, the results for the machine learning algorithm used for detecting phishing websites for the decision tree is 97.01% accurate with NLP feature and 82.47% accurate with World Vector.

VII. Conclusion

Web application attacks are managed by the various users. Among all the other approaches presented in the related work, it has been identified that the Decision tree is an efficient way to prevent the new kinds of attacks and protect the security and information of user. A methodology has been proposed to predict the phishing vulnerabilities using decision tree algorithm is presented. It has been analyzed and executed with the various users' credentials. The results suggested that the performance of the decision tree algorithm used in preventing Phishing Attack. Decision Tree approach is presented in this paper is efficient and provides a good mechanism to identifying attack patterns. This methodology has been executed and concluded with various remarks in the conducted results.

References

[1] Abdul Razzaq, Ali Hur, Sidra Shahbaz, Muddassar Masood, H Farooq Ahmad- "Critical Analysis on Web Application Firewall Solutions", Issue, 2013
[2] Abdulrahman Alzahrani, Ali Alqazzaz, Huirong Fu, Nabil Almashfi, Ye Zhu- "Web Application Security Tools Analysis", Issue, 2017.
[3] Dennis Appelt, Cu D. Nguyen, Lionel Briand- "Behind an Application Firewall, Are We Safe from SQL Injection Attacks?", Issue, 2015

DOI : 10.46528/JK.2020.V10I06N03.027

ISSN: 2278-4632
Vol-10 Issue-6 No. 3 June 2020

[4] Dennis Appelt, Annibale Panichella, Lionel Briand- "Automatically Repairing Web Application Firewalls Based on Successful SQL Injection Attacks", Issue, 2017
[5] Sandeep Kumar1, Renuka Mahajan2, Naresh Kumar3, Sunil Kumar Khatri- "A Study on Web Application Security and Detecting Security Vulnerabilities", Issue, 2017
[6] Dennis Appelt, Cu D. Nguyen, Annibale Panichella, and Lionel C. Briand, *Fellow, IEEE*- "A Machine-Learning-Driven Evolutionary Approach for Testing Web Application Firewalls", Issue, 2018
[7] Ram Basnet, Srinivas Mukkamala, Andrew H. Sung- "Detection of Phishing Attacks: A Machine Learning Approach" Issue, 2008
[8] Yasin Sönmez, Türker Tuncer, Hüseyin Gökal, Engin Avci- "Phishing Web Sites Features Classification Based on Extreme Learning Machine" Issue, 2018
[9] Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, Banu Diri -" Machine learning based phishing detection from URLs", Issue, 2018
[10] A. D. Brucker, L. Brugger, P. Kearney, and B. Wolff- "Verified firewall ¨ policytransformations for test case generation", Issue 2010.
[11] D. Appelt, N. Alshahwan, and L. Briand- "Assessing the impact of firewalls and database proxies on SQL injection testing", Issue 2013.
[12] D. Gupta, J. Bau, E. Bursztein, and J. Mitchell- "State of the art: Automated black-box web application vulnerability testing", Issue 2010.
[13] W. Halfond, J. Viegas, and A. Orso- "A classification of sql-injection attacks and countermeasures", Issue 2006.
[14] Y.F. Li, P. K. Das, and D. L. Dowe- "Two decades of web application testing: A survey of recent advances", Issue 2014.
[15] Tajpour, Maslin Masrom, Mohammad Zaman Heydari, Atefeh, and Suhaimi Ibrahim- "SQL injection detection and prevention tools assessment", Issue 2010.

DOI : 10.46528/JK.2020.V10I06N03.027