A MACHINE LEARNING BASED MECHANISM TO AUTO-DETECT CYBER ATTACKS

Ganesh Sunil Bayas*1, Charudatta Kapade*2, Diksha Kadam*3, Raviraj Nagul*4, Jayashree C. Pasalkar*5

Student, Information Technology, AISSMS Institute of Information Technology, Pune, India.
*5Assistant Professor Information Technology, AISSMS Institute of Information Technology, Pune, India.

ABSTRACT

Sharing information over the internet through multiple platforms and web applications has become quite common and essential in the recent times. These web applications that accepts such critical information collected from users store this information in databases. Due to being accessible through internet these applications and the databases which is connected to them are prone to all kinds of information security threats. The threats having various kinds of attack such as Cross Side Scripting (CSS), Denial of Service(DOS), and SQL injection attacks. Through this kind of attack the intruder can easily steal a confidential information and tamper it. Hence it could have extreme damaging effects on a business or organisations. From the top ten vulnerabilities SQL injection attacks is one of the attacks which may cause or result in confidential data being deleted or web applications being defaced and so on. The effects could range from momentary loss, decrease in company’s stock market value, leaking confidential business information or any combination of these. In this project we have build a cyber security system using Naive Bayes algorithm from ensemble machine learning approaches to classify and detect SQL injection attacks and DOS attacks.

Keywords: machine learning, Denial of Service(DOS), SQL injection attacks, Naive Bayes, cyber Security, Deep learning.

I. INTRODUCTION

Usually most of the applications are web based applications that we use in our day to day life. As organisations in order to achieve increase in their exposure they choose to make their applications available on internet. So that their web applications is accessed by various users globally. And hence while being exposed to internet it increases the chances security of threats occurrences that come along with uncontrolled access. As there is a huge increases in the usage of internet we are likely comfortable in performing various kinds of transactions online each and every single data entered while transaction on various websites in stored in some kind of database. These relational databases can be merged with a language called structured Query language. When SQL is used for launching attacks on the databases in order to manipulate the users data is a form of web hacking technique known as SQL injection attacks can cause a severe security threat to the organisations. As it may result in deletion of confidential data, website being hacked, data loss or stolen, unauthorized access to system or account and may finally lead to individual machine or entire network breakdown. Previously support vector machine (SVM) algorithms was used for classification and prediction of SQL injection attack. In our purpose algorithm we use Naive Bayes algorithm for it has highest accuracy rate than SVM i.e 99.5% as SVM has only 96.47%. we use machine learning algorithm to detect and prevent various cyber security threats which are being increasingly deliberately tremendous research work has been done on using various machine learning algorithms to detect SQL injection attacks. In our previous research we saw that the cost train SVM for large train set was the clear drawback as compared to Naive Bayes it is very simple and easy to understand.

II. PROPOSED METHODOLOGY

In fig.1.The system working is explained in which the principal parts of function are represented by blocks connected by lines that show the relationship of blocks and working of system. System contains five blocks name Security System, Behavioural Analysis, Database, ML Algorithm, Action. In First phase user Access the Security System then he select the option Activates Behavioural analysis and Runs the ML Algorithm if any malware of Network intrusion or any unusual activities is found the System Takes action and alert the database and runs ml algorithm. This is working of system.
III. IMPLEMENTATION

Figure shows a box-graph plot of comparisons of performance of various machine learning algorithms after applying them to various datasets. Naïve bayes algorithm has highest performance than other machine learning algorithms. The algorithm fall under Ensemble learning category.

Hence Naïve Bayes Classifier was chosen to implement SQL Injection detection on our dataset. Naïve bayes Classifier has been implemented from the ensemble part of Scikit-Learn library in Python. Parameter tuning is of high significance in ensemble learning algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>SQL injection accuracy</th>
<th>Denial of service accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB classifier</td>
<td>99.5%</td>
<td>99.2%</td>
</tr>
</tbody>
</table>

TABLE I. Performance measures play an important roles in machine learning. They are not only used as the criteria to evaluate learning algorithms, but also used as the problem solving technique to construct learning models. The accuracy of Naïve bayes classifier to detect SQL injection attacks stands at 99.5% and for Denial of service attacks at 99.2% respectively.
Figure shows a Detection of website attacks and Server attacks. The Packets Are analyse By detection Engine the packets which contains malicious data in frames and Malicious HTTP request are examine by rules if it contains Vulnerable url or any sql query which is crafted by attacker it’s get trigger and get logged in Database. Our Systems is also capable of detecting port scan on server, or any bruteforce on ftp,ssh service or any UDP Flood DDos Attacks.

IV. DESCRIPTION ON DATASETS

A. Defense Advance Research Project Agency (DARPA)-

The system prior used DARPA data sets 1999 its major drawback was that it was time consuming and was difficult to obtain a representative data set. These data sets was so huge that the researchers too worked of other subsets.

B. Netflow-

Later DARPA the netflow dataset came into existence which were used in prior System. Netflow does not have the important feature such as tcp dump. It was introduced as a router feature by the CISCO. Usually the netflow contains three main components which are

- A netflow exporter
- Netflow Collector
- Analysis console.

Its feature were limited only to flow the information which was generated by higher end routers.

C. SQL Injection Dataset-

The dataset for SQL Injections has been created from a tool named Libinjection. Libinjection is an open source tool that is used for penetration testing of web applications. It passes SQL Injections as payload to web applications and analyses if the application is vulnerable to SQL Injection attack. By the use of this tool, all the payloads generated by libinjection were captured for a particular instance and a dataset consisting of all these payloads is used as the SQL Injection dataset. This dataset contains around six thousand SQL Injections of all the three types, that are, Union Based, Error Based and Blind SQL Injections.

V. CONCLUSION AND FUTURE WORK

In this paper we present a user- centric machine learning system which helps to carry out the SQL injection attacks on the web applications and eliminate the risk of system damage or breakdown. We made sure that all the drawbacks which were present in prior existing systems were overcome and also increases the accuracy rate via using Naïve Bayes Classifier. Not only SQL injection attacks but also the other types of bot attacks can be detected through this systems. In future we intend to cover other network attacks in our system.

VI. REFERENCES


[2] Sumouli Choudhary and Anirban Bhowal. “Comparative Analysis of Machine Algorithms along with...
Classifiers for Network Intrusion Detection”.


