Suspicious Activity Detection on E-Commerce Application

Meghana C
Dept. of Information Science and Engineering
Acharya Institute of Technology
Bengaluru, India

Chaitra B
Dept. of Information Science and Engineering
Acharya Institute of Technology
Bengaluru, India

Abstract—With the quick utilization of web, web applications and internet web applications are rising as well. It is often the case that e-commerce applications face various suspicious activity such as slow loading of pages, unexpected pages displaying and SQL injection vulnerabilities. This paper addresses one such suspicious activity that is slow loading of pages resulting from imposing too many requests on the server. The results will be classified as low, medium and high suspicious activity by executing LCS and SVM algorithm.

Keywords—LCS; Suspicious Activity;

I. INTRODUCTION

Web applications are PC programs permitting site guests to submit and retrieve information to/from database over the web utilizing internet browser. Web applications play a crucial role in enhancing customer support. The main usage of application is that they can be accessed at any time and from any place. Slow loading of pages is a non-intrusive web attack made to bring down the focused site or to back it off by flooding the server or application with counterfeit traffic [12]. The traffic is initiated from imposing too many requests on the server. To manage the traffic on the server each user activity has to be observed and managed. This user activity is tracked by graphical link tracking concept and user activity is stored in habitat file [16].

II. OBJECTIVE

The primary thought process of this project is to overcome the suspicious activity in e-commerce application. One of the way used to achieve this is by tracking each button, click, and the navigation patterns (pages) of the users for each session and then obtain the habitat of the user. Then classify the users for a given set of sessions into Low, Medium and High Suspicious. Block the user for a certain period of time to avoid any kind of operations.

III. LITERATURE SURVEY

Ming D. Wan et al [1] suggested that, intruders on the Internet can be distinguished by taking a look at the resemblance of two thumbprints. The measurement of LSS is an estimation of likeness between the two groupings. The LSS which is known as Longest Similar Subsequence issue is a theory of the remarkable problem, Longest Common Subsequence (LCS). In intrusion detection, it may be important to contrast two groupings of thumbprint to check they are comparable.

Ibrahim Salim et al [6] describes that DoS attack allows the intruders to gain the services of network thus stopping the access to legitimate users. To overcome this, it is fundamental to design a system to recognize intrusion called IDS. It is a kind of security programming that works as a system security framework to shield the PC framework from malicious attacks. As the level of information being shared from one network to other is increasing, the IDS help in identifying the interruptions effectively. Data mining is a productive tool applied to layout of IDS and keep the gigantic system information from the interlopers. Exceptions are designs in information that don't match to a well-defined normal behavior.

A. K. Tripathy et al [10] recommended that service option for automatic service composition with customer's prerequisites oriented service selection turns out to be increasingly exceptional. The current planning and selection algorithms are generally intended for service disclosure. Further, as far as anyone is concerned, there are just a couple of works that consolidate end-client prerequisites into service composition.

M Khan et al [11] performed analysis and proposed review on various other different outlier detection strategies from the perspective of data mining. Current studies in data mining is center for discovering unique patterns from huge datasets and utilizing it for making decision. However, discovering outliers and exceptions did not get lot of consideration in data mining field as different topics got.

Cheng Zong I et al [12], Conducted research where in the LCS issue has been applied to IDS, E-business and Bioinformatics and so forth. This paper suggests an extended longest common subsequence called LCS problem, designs an equal calculation to tackle LCS issue on SMP machine by partition and vanquish technique.

IV. METHODOLOGY

To implement the methodology, first an e-commerce application is developed on the local host. To test our approach, we initially created 20 users and user details were stored in the habitat file. All the transactions that are performed inside the website is been tracked via Graphical link tracking.
In this project, focus is on imposing max no of requests on a server and then classifying those requests as low, medium and high suspicious. When an intruder logs into the website, the transactions performed by him will be saved in habitat file. This can be monitored by the admin by running the LCS algorithm.

4. Find the maximum separable width for the prediction data.
5. Lowest distance plane width is assigned to class.

C. Weight Computation

Each transactions performed by the user creates a unique pattern along with User ID and Session ID which is in turn stored in the habitat file. This habitat file contains user ID, session ID, unique patterns of users, frequency of unique patterns. The actions (number of pages visited) performed by each user is considered as a string L [16]. The length of the string is divided and LCS value is obtained. LCS algorithm is used for finding longest subsequence common to all sequences in a set of sequence [11]. LCS value contains different users and actions performed by them. The average unique pattern is calculated by Count of Unique Patterns

\[
W_q = \frac{f_j}{f_j + 0.5 + 1.5n_{sj}} \frac{\log (N + 0.5)}{\log (N + 1)}
\]

D. Implementation

1. Creating a Product Shopping website with around say 5 products where user will be able to perform user registration, login, product list & product buys. The Product Shopping website will make use of the following tables
   a. Customers’ Information Table: Table of client’s data stores data of all clients like login Id, Secret key etc.
   b. Product Information Table: Stores the data pretty much all the items sold by the site.
   c. Category Table: Stores all the category Id and the respective item’s category name.
   d. Order Information Table: These are utilized to store data about the requests the client has positioned. It stores the orders, Product Id, amount and so on.[13]
2. The actions performed by the user are captured in a system known as Habitat file.
3. The Admin can log in and analyze the algorithm and find out within a given period of transactions the pattern of navigation and most predictive user responsible for suspicious activity.
4. Clusters of such suspicious activity users are formed by making use of LCS algorithm.
5. Clusters of such suspicious activity users are formed by making use of LCS algorithm.
V. RESULTS

The outcome of the implementation is: when only SVM algorithm is used to detect suspicious activity, only the latest session activity will be shown. Whereas LCS algorithm resulted in displaying the activity across the session.

VI. CONCLUSION

The customer will be able to register into the application and all the user activities are stored in the database file called habitat file. The habitat file will have the tracking based on action name, action type, time and date of session along with user id as well as session id. The Admin will be able to dynamically determine the Suspicious activity by executing the LCS and SVM algorithm and then the user will be classified as high, medium and low suspicious if the repeated actions are performed based on the patterns. Such users can be blocked by the admin and further login is not possible. This can be extended further to detect SQL Injection and Cross Site Scripting attack.

ACKNOWLEDGMENT

I extend my gratitude to Prof. Marigowda C K, Head of the department, ISE Branch and Dr. Surekha K B, Associate Professor of Acharya Institute of Technology for their guidance and support.

REFERENCES


