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.

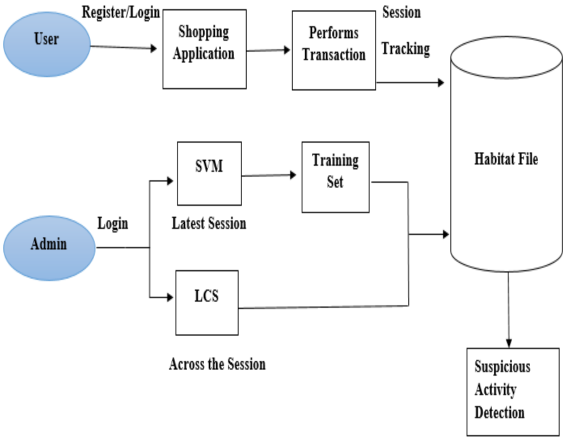


Fig. 1. Process Flow Diagram

A. LCS Algorithm

The LCS algorithm is described as follows: Let us consider 2 strings X and Y. Consider x as the length of string X and y as the length of string Y. The longest succession of characters which should basically be in a bordering way that can be found in two strings X and Y is known the LCS.[6]

1) Consider the sequences namely S1 and S2
2) The length of S1 and length of S2 is computed
3) Find the maximum length of S1 and S2
4) Construct initial matrix with 1st row and 1st column with zeros.
5) If the value of the sequence alphabet is not there then maximum on top and left is taken
6) [2][3] If the value matches then diagonal value is incremented by a value of 1.

B. SVM Based Suspicious Detection

This algorithm is described as a discriminative classifier formally portrayed as an isolating hyper plane.[4] When the training data is given, the figuring gives a perfect hyper plane which characterizes new models. In 2D space, a plane is divided in two sectors by a hyper plane where in each class lay in either side.[5]

The study of the hyper plane in Linear SVM is carried out by changing the issue using some linear algebra. Kernel plays an important role here. The condition to predict new input using the dot product between the input (x) and each help vector (x_i) in Linear Kernal is:

1. Attribute values for attribute names for which class has to be predicted.
2. Training data is plotted on a plane.
3. Construct a hyper plane line in order to separate data points.
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 $\frac{\text{Unique Patterns}}{\text{Number of unique}}$

$$\boxed{W_{ij} = \frac{f_{ij}}{f_{ij} + 0.5 + \frac{1.5 n_{sj}}{ns_{avg}}} \quad \frac{\log\left(\frac{N+0.5}{M_i}\right)}{\log(N+1)}}$$

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.

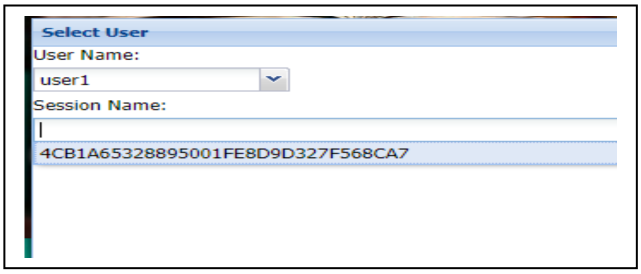
The Product Shopping website will make use of the following tables

   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.

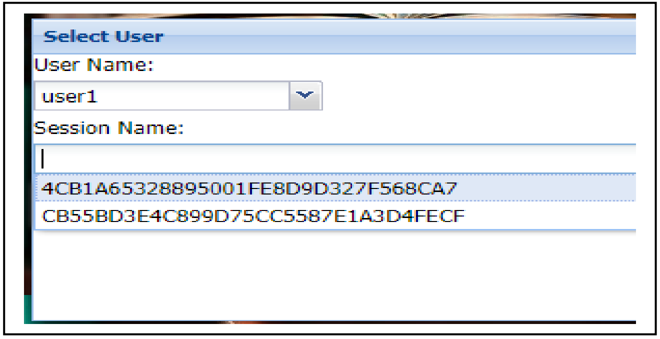
V. RESULTS

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.



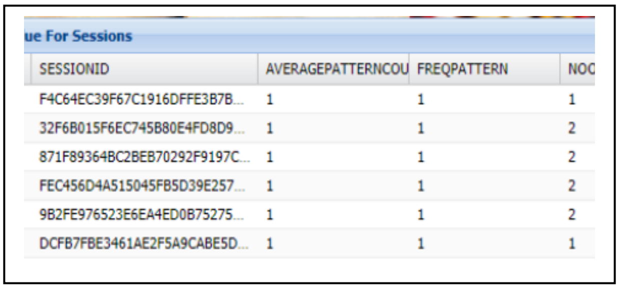
Select User
User Name:
user1
Session Name:
|
4CB1A65328895001FE8D9D327F568CA7

Fig. 2. SVM displaying latest Session Activity



Select User
User Name:
user1
Session Name:
|
4CB1A65328895001FE8D9D327F568CA7
CB55BD3E4C899D75CC5587E1A3D4FECF

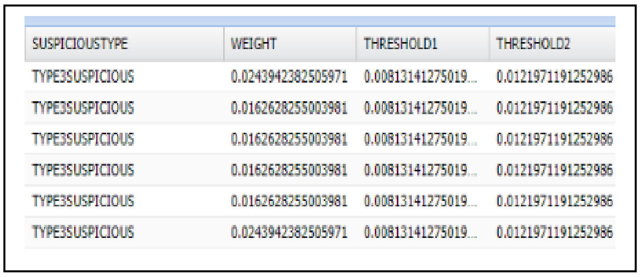
Fig. 3. Across Session activity by LCS



ue For Sessions

SESSIONID	AVERAGEPATTERNCOU	FREQPATTERN	NOC
F4C64EC39F67C1916DFFE3B7B...	1	1	1
32F6B015F6EC745B80E4FD8D9...	1	1	2
871F89364BC2BEB70292F9197C...	1	1	2
FEC456D4A515045FB5D39E257...	1	1	2
9B2FE976523E6EA4ED0B75275...	1	1	2
DCFB7FBE3461AE2F5A9CABE5D...	1	1	1

Fig. 4. LCS o/p displaying the users Session ID and pattern count



SUSPICIOUSTYPE	WEIGHT	THRESHOLD1	THRESHOLD2
TYPE3SUSPICIOUS	0.0243942382505971	0.00813141275019...	0.0121971191252986
TYPE3SUSPICIOUS	0.0162628255003981	0.00813141275019...	0.0121971191252986
TYPE3SUSPICIOUS	0.0162628255003981	0.00813141275019...	0.0121971191252986
TYPE3SUSPICIOUS	0.0162628255003981	0.00813141275019...	0.0121971191252986
TYPE3SUSPICIOUS	0.0162628255003981	0.00813141275019...	0.0121971191252986
TYPE3SUSPICIOUS	0.0243942382505971	0.00813141275019...	0.0121971191252986

Fig. 5. LCS o/p displaying Type of Suspicious and Threshold value

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

[1] Ming D. Wan, Shou-Hsuan Stephen Huang, and Jianhua Yang, " Finding the Longest Similar Subsequence of Thumbprints for Intrusion Detection" , Proceedings of 20th Inter Advanced Information Network(AINA) 2017.

[2] S. Çalışır, R. Atay, M. K. Pehlivanoğlu and N. Duru, "Intrusion Detection Using Machine Learning and Deep Learning Techniques," 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, 2019, pp. 656-660.

[3] L. Aslanyan, "LCS algorithm with vector-markers," 2017 Computer Science and Information Technologies (CSIT), Yerevan, 2017, pp. 92-96.

[4] N. Putpuek, N. Cooharojananone and S. Satoh, "A modification of retake detection using simple signature and LCS algorithm," 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Kanazawa, 2017, pp. 257-261.

[5] V. Justin, N. Marathe and N. Dongre, "Hybrid IDS using SVM classifier for detecting DoS attack in MANET application," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, 2017, pp. 775-778.

[6] Ibrahim Salim, T.A.Razzack,"A study on IDS for Preventing denial of service attack using outliers techniques", 2nd IEEE international conference on Engineering and technology, March 2016.

[7] F. Mira, A. Brown and W. Huang, "Novel malware detection methods by using LCS and LCSS," 2016 22nd International Conference on Automation and Computing (ICAC), Colchester, 2016, pp. 554-559.

[8] R. M. Pandurang and D. C. Karia, "Impact analysis of preventing cross site scripting and SQL injection attacks on web application," 2015 IEEE Bombay Section Symposium (IBSS), Mumbai, 2015, pp. 1-5.

[9] V. Anitha, S. A. Lakshmi, M. Revathi and K. Selvi, "Detecting various SQL Injection vulnerabilities using String Matching and LCS method," 2014 Sixth International Conference on Advanced Computing (ICoAC), Chennai, 2014, pp. 237-241.

[10] A. K. Tripathy, M. R. Patra, M. A. Khan, H. Fatima and P. Swain, "Dynamic Web Service Composition with QoS Clustering" 2014 IEEE International Conference on Web Services, Anchorage, AK, 2014, pp. 678-679.

[11] M.Khan , S.K.Pradhan, M.A.Khaleel, "Outlier Detection for Business Intelligence using data mining techniques", International journal of Computer Applications (0975 -8887), Volume 106- No. 2, November 2014.

[12] Cheng ZhongI,Guo-Ling Chen, Jia-Hua He, " Parallel Computing for the Longest Common Subsequences in Network Intrusion Detection System", Proceedings of the Third International Conference on Machine Leaming and Cybermetics, Shanghai, 26-29 August 2014

[13] R. D. Rubi and L. Arockiam, "Positional_LCS: A position based algorithm to find Longest Common Subsequence (LCS) in Sequence Database (SDB)," 2012 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2012, pp. 1-4.

[14] K. Haizhou, "A Research of Attacking Access Controls Algorithms of Web Application," 2012 International Conference on Industrial Control and Electronics Engineering, Xi'an, 2012, pp. 679-681

[15] Y. S. Sneha, G. Mahadevan and M. M. Prakash, "An online recommendation system based on web usage mining and Semantic Web using LCS Algorithm," 2011 3rd International Conference on Electronics Computer Technology, Kanyakumari, 2011, pp. 223-226.

[16] R. Priyadarshini, D. Jagadiswaree, A. Fareedha and M. Janarthanan, "A cross platform intrusion detection system using inter server communication technique," 2011 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, Tamil Nadu, 2011, pp. 1259-1264.

[17] Z. Xu and Y. Yu, "LCS Algorithm Based on Similarity of Gene Sequences," 2011 Third Pacific-Asia Conference on Circuits, Communications and System (PACCS), Wuhan, 2011, pp. 1-3.

[18] A. Yamada, H. Masanori and Y. Miyake, "Web Tracking Site Detection Based on Temporal Link Analysis," 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, Perth, WA, 2010, pp. 626-631.

[19] Niksefat, Salman &Ahaniha, Mohammad &Sadeghiyan, Babak&Shajari, Mehdi. (2010). Toward Specification-Based Intrusion Detection for Web Applications. 6307. 510-511. 10.1007/978-3-642-15512-3_3

.