International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

An Integrated Prediction of SQL Injection using Random Forest

S.Sivamohan, L. Kali Prasad, Vigneshwararaj. S, A.I. Vishal

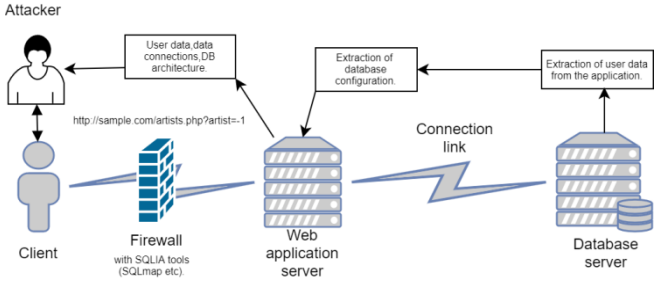
ABSTRACT- SQL injection is one of the cyber attack used by the attackers to penetrate into the web application database. This attack is considered to be the top ten threats and is also declared by "Open Web Application Security Project". The importance of the injection detection is that even a young person can initiate this attack from any place and also no prior knowledge is required as there are existing tools available extensively. This attack works in the way by inserting a malicious code or logic in the authentication page and this compromises the system to return true in the condition while checking the data with the data present in the database. Actually, this malicious code breaks the format of string to a logic based function as in default all the data that are inputted by the user is written in a string format. We are using Random Forest algorithm to detect the injection attack.

Keywords : - SQLIA (SQL Injection Attack),CIA (Central Intelligence Agency)..

I. INTRODUCTION

SQL Injection is the type of cyber attack performed by the attackers to penetrate and to retrieve the data stored in the database. Now, in recent days there are lots and lots of web applications that are being used world wide. Injection attack is specifically used to retrieve the data from the web applications. Injection attack uses some malicious code to enter into the server. SQL injection attack can be used in two ways. One of the way is by the use of POST method and another way is by the use of GET method. In the GET method the malicious code is written on the uniform resource locator directly as this method provides the data on the URL. Coming to the POST method, the attack that is done by using this method is on the user login page. The malicious login is written in the user credentials rather than the original data that is to be inputted. Basically how this web application interaction with the user works is that, it first asks the user for the credentials that are to given, it includes the user name and the password given by the user as shown in figure[1]. After completion of the the authentication the user will be directed to the server page and in which the user can access the data and also can modify the data. The data that has been provided by the user is taken as string input and also the data that is stored in the database server is also in the string format. This works in the way such that the given client data and the data stored in the database are compared with each other to return. This is the simple way of accessing the web application. Coming to the injection attack, how it works is that it uses its required logic to compromise the authentication. The user data is taken as string input and it uses its logic in the place of string input to manipulate the server to provide access to its entry as it compares with the data stored in the database only if the data is considered to be as in the string format, it breaks the format as its malicious logic. Also this attack is similar to that of OR gate where it returns true even if any of the credentials written is false. Normally OR gate works in the condition to satisfy the system if one condition among two is written correctly.

Saying so about the SQL injection attack and its working methodology, this attack can be effective only on SQL(structured query language) database. In this modern era of emerging technologies with more advances there are new entries of ideas eliminating the previously used technique with a better feature. All the launch entries are not taken into account due to consistent replacement but SQL databases hold its significant role in relational database. With the boom in advancement of tech the users switch to web based services and applications. The web services and applications integrates with a database for storing the data. This database is vulnerable for attacks of different sort to retrieve users data. In this paper, we propose SQL injection detection model with random forest algorithm with the signified query analysis from the log captured from the web server. This detection model segregates the quality of the query with comparison to the training data set which has a quite simplistic and efficient process to detect SQLIA(SQL injection attack).



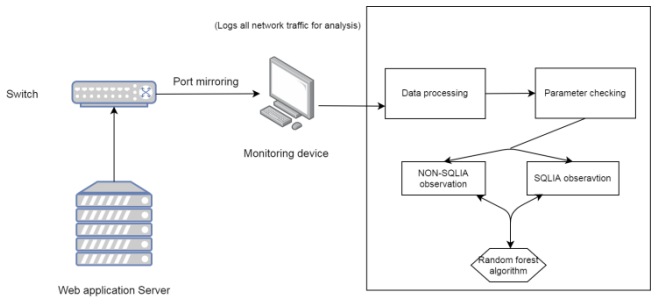
Figure[1]

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Mr. Sivamohan S*, Assistant professor of Information technology in SRM IST,Chennai,Tamil Nadu,India. Email: sivamohs@srmist.edu.in.
L. Kali Prasad, department of Information Technology,SRM IST,Chennai,Tamil Nadu,India. Email: lkpkali@gmail.com
Vigneshwararaj. S, department of Information Technology,SRM IST,Chennai,Tamil Nadu,India. Email: vigneshwarraaj21@gmail.com
A.I. Vishal department of Information Technology,SRM IST,Chennai,Tamil Nadu,India.Email: aivishal2020@gmail.com.

II. LITERATURE SURVEY

From the authors of[5] their work completely worked for the intrusion detection system using the defined algorithms for classifying the type of cyber-attacks took place. This system monitors continuous change in the network state and user activity. All of which needs more constant monitoring all over the working system with repeated cross checking in the classification of the behavior and this makes the system much complex in working. It has a difficulty in distinguishing the network, normal behavior and any internal error will impact the system to generate a false output overall. The data sets like KDDcup 99 data set are outdated but collective for the cyber security improving community[5]. From the authors [1] we can understand the normal basis and every perspective of an attack to happen. Everything from the CIA triad to deployment related security functions related to web based attacks are noted but they tend miss over in the input validation and code injected into applications for attackers benefits[1].

.



Figure[2]

III. EXISTING SYSTEM

This system works on the basis of specifications. The drawback in the existing system is that it cannot detect all the types of injection attacks. The enhancement provided in the proposed system is that it can be able to detect all types of injection attacks. Also in this system it cannot have a direct access to the database.

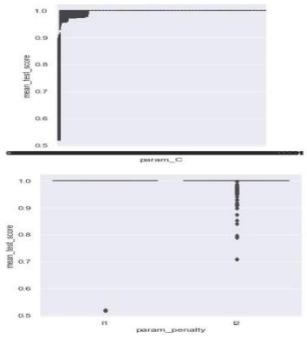
The database accessibility is restricted by user authentication through login service. All the applications commands and scripts are saved as logs. The logs are analyzed for any kind of malicious scripts or commands for the SQL injection vulnerability. The process of log entry and detection is inefficient in the current system.

IV. IMPLEMENTATION

In this proposed system we are using random forest algorithm to detect the SQL injection attack. The main advantage of using this random forest algorithm in the detection of attack is that it provides high accuracy when compared to that of previously used systems. Also this works with the help of two data sets that are provided with the features. Non SQL injection observation and SQL injection observation are the two data sets provided as shown in the figure[2]. The malicious logic that are written to penetrate into the database system comes under the SQL injection observation. This observation captures the track of the data that is considered to be malicious. Non SQL observations includes the data which are not considered to be harmful to the system. In the data sets provided to predict the accuracy of detection there are five attacks of SQL injection and also provided with the label to monitor the entry. Also the data sets consist of codes like 1 and 0 in which 1 can be used to refer the attack that has been made to penetrate into the system and 0 can be defined to notify that no attacks are made. The data sets are saved in the excel format and is imported with the help of the pandas package.The algorithm used in this system is imported with the help of sklearn package and also it comes under ensemble learning. Also train test split package is imported from the sklearn which is used to train and test the data. This works in the way such that the features of the data sets are divided into halves and the input data is taken as the data for train x and the output of those features are taken in the train y. The test x and test y are used to test the features to predict the desired output.

V. RESULT

Best Model Hyper Parameters: {'C': 0.0171, 'penalty': 'l1'}
Test set accuracy = 1.0.



Figure[3]

VI. FUTURE SCOPE

As the technology keeps on growing vast, there is always a way for more problems to arise from the other side of ethics. The state of depending the attacks and malicious services keeps up more with the advanced technology and techniques which is equivalently crossing up with the attackers who are also having the same access to the technology and techniques. Everything said with more advancement there should be more complex methods for protection and prevention. For keeping up with the trend we have planned to give more futuristic features as updates which can enhance this prediction model to work more efficiently giving more accurate results. One of the feature which can give further accuracy in prediction is having a separate method to log or note all the unclassified observations and these unclassified observations are analyzed for activity taken on.These observation need some manual overview from the security team for further refinement in results and the look for more malicious activities like backdoor etc.

VII. CONCLUSION

This system helps to detect the SQLIA attack efficiently as we are using required data sets for the detection of attack and also it provides the confidentiality of the user data that is stored. The data sets provided helps us to identify the malicious entry of unknown data by the users. This helps the system to avoid loss of data records.

REFERENCE

1. D. Mitropoulos, P. Louridas, M. Polychronakis, and A. D. Keromytis, ``Defending against Web application attacks: Approaches, challenges and implications," IEEE Trans. Depend. Sec. Comput., vol. 16, no. 2, pp. 188203, Mar./Apr. 2017.
2. A. Sarhan, S. A. Farhan, and F. M. Al-Harby, ``Understanding and discovering SQL injection vulnerabilities," in Proc. Int. Conf. Appl. Hum. Factors Ergonom., 2017, pp.
3. 10631075.
4. D. Das, U. Sharma, and D. K. Bhattacharyya, ``Defeating SQL injection attack in authentication security: An experimental study," Int. J. Inf. Secur., vol. 18, no. 1, pp.
5. 122, 2017.
6. B J. Santhosh Kumar and P. P. Anaswara, ``Vulnerability detection and prevention of SQL injection," Int. J. Eng. Technol., vol. 7, no. 2.31, pp. 1618, 2018.
7. [algorithm in large IOT", Computer Communications, Volume 148, 15 December 2019, Pages 107-114.
8. R.Mythili, Revathi Venkataraman, T.Sai Raj,"An attribute-based lightweight cloud data access control using hypergraph structure", The Journal of Supercomputing(JoS),Published online: 02 Jan 2020 DOI: 10.1007/s11227-019-03119-7.
9. S.Sivamohan, Liza.M.K, R.Veeramani, Krishnaveni.S, Jothi.B, "Data Mining Techniques for DDOS Attack in Cloud Computing", IJCTA Interntional Scoience Press, Pg: 149-156
10. S Pandiaraj, Aishwarya, Surbhi, Alisha Minj, Priyanshu Singh, "Enabling Cloud Database Security Using Third Party Auditor", International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-4, April, 2019.
11. R.Veeramani,Dr.R.Madhan Mohan, "Iot Based Speech Recognition Controlled Car using Arduino", International Journal of Engineering and Advanced Technology,Volume-9 Issue-1, October 2019.
12. T.H. Feiroz khan, N.Noor Alleema, Narendra Yadav, Sameer Mishra, Anshuman Shahi "Text Document Clustering using K-Means and Dbscan by using Machine Learning",International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019.
13. S.Babeetha, B. Muruganantham, S. Ganesh Kumar, A. Murugan, "An enhanced kernel weighted collaborative recommended system to alleviate sparsity", International Journal of Electrical and Computer Engineering (IJECE), Volume 10, February 2020, Page No. 447-454.
14. [13] Kavitha.R ,K.Malathi,"Recognition and Classification of Diabetic Retinopathy utilizing Digital Fundus Image with Hybrid Algorithms", October 2019,International Journal of Emgineering & Advanced Technology(IJEAT), Volume 9, Isssue 1, 109-122.
15. T.Chandraleka,Jayaraj R, " Hand Gesture Robot Car using ADXL 335 ", International Journal of Engineering and Advanced Technology (IJEAT)', Volume-8 Issue-4, Nov 2019
16. H.Sangeetha,S.Abinayaa, "Smart Irrigation Systems using Sensors and GSM" in 'International Journal of Recent Technology and Engineering (IJRTE)', Volume-8 Issue-1, May 2019. Page No.:884-886.
17. B.Sathya Bama,,Y.Bevish Jinila, "Attacks in Wireless sensor networks- A Research" ,International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8, Issue-9S2, July 2019.
18. Vellingiri, J., S. Kaliraj, S. Satheeshkumar and T. Parthiban , "A Novel Approach for User Navigation Pattern Discovery and Analysis for Web Usage Mining", Journal of Computer Science 2015, vol 11 (2): Page no 372.382.

AUTHORS PROFILE

Mr. Sivamohan S is an Assistant Professor (Selection Grade) in Department of Information Technology, SRM IST, Chennai, Tamil Nadu, India.

L. Kali Prasad is currently pursuing bachelors of technology in information technology from SRM IST,Chennai,Tamil Nadu,India.

Vigneshwararaj. S is currently pursuing bachelors of technology in information technology from SRM IST,Chennai,Tamil Nadu,India.

A.I. Vishal is currently pursuing bachelors of technology in information technology from SRM IST,Chennai,Tamil Nadu,India.