IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 4, April 2019

Intrusion Detection using Machine Learning and log analysis

Miss. Manali Raka[1], Miss. Shrushti Shah[2], Miss. Sayali Gunale[3],

Miss. Renuka Tanksale[4], Prof. Mr U.K.Raut[5]

Comp Dept, Maharashtra Institute of Technology, Pune, India[1,2,3,4]

Professor, Comp Dept, Maharashtra Institute of Technology, Pune, India[5]

Abstract: Web-based applications has gained universal acceptance in every field of lives, including social, commercial, government, and academic communities. Even with the recent emergence of technology, most of applications are accessed and controlled through web interfaces this work proposes a multistage log analysis architecture, which combines both pattern matching and machine learning methods. It makes use of logs generated by the application during attacks to effectively detect attacks and to help preventing future attacks. The architecture is explained in detail with a proof-of-concept prototype is implemented using pattern matching and Bayes Net for machine learning. Experiment outcomes show that the two-stage system has combined the advantages of both systems, and has improved the detection accuracy. The proposed work is significant in advancing web securities, while the multi-stage log analysis concept would be highly applicable to many intrusion detection applications. In proposed system we are use Intrusion detection technology on net banking application to detect various types of attack that affect net banking application.

Keywords: SQL injection, Machine Learning, Web Security

I. INTRODUCTION

An Intrusion Detection System (IDS) monitors the network traffic looking for suspicious activity, which could represent an attack or unauthorized access. In the recent year demand for huge data storage and high processing speed has led to the era of cloud computing, and this systems and applications are accessed through the Internet web. In Proposed system naive baiyes algorithm are used. to protect web-based systems and net banking applications, strong, effective security mechanisms must be placed. Log analysis is to detect when an intruder attacked the system is what steps were performed during the attack and, how it happened, during the attack. The manual log analysis is no longer possible for real-time log analysis, as the number of logs has quickly increased. Manual analysis also involves more cost and much more time.. Net banking application is the area where we found various types of attack that affects whole system. In this application we use intrusion detection on net banking application. To detect various detect that done on net banking application.

II. LITERATURE SURVEY

Title: Detection model for SQL injection attack: An approach for preventing a web application from the SQL injection attack
Author: G. Buja, K. Bin Abd Jalil, F. Bt Hj Mohd Ali, & T.F.A. Rahman
Description: Since the past 20 years the uses of web in daily life is increasing and becoming trend now. As the use of the web is increasing, the use of web application is also increasing. Apparently most of the web application exists up to today have some vulnerability that could be exploited by unauthorized person. Some of well-known web application vulnerabilities are Structured Query Language (SQL) Injection, Cross-Site Scripting (XSS) and Cross-Site Request Forgery (CSRF). By compromising with these web application vulnerabilities, the system cracker can gain information about the user and lead to the reputation of the respective organization. Usually the developers of web applications did not realize that their web applications have vulnerabilities. They only realize them when there is an attack or manipulation of their code by someone. This is normal as in a web application, there are thousands of lines of code, therefore it is not easy to detect if there are some loopholes. Nowadays as the hacking tools and hacking tutorials are easier to get, lots of new hackers are born. Even though SQL injection is very easy to protect against, there are still large numbers of the system on the internet are vulnerable to this type of attack because there will be a few subtle condition that can go undetected. Therefore, in this paper we propose a detection model for detecting and recognizing the web vulnerability which is; SQL Injection based on the defined and identified criteria. In addition, the proposed detection model will be able to generate a report regarding the vulnerability level of the web application. As the

consequence, the proposed detection model should be able to decrease the possibility of the SQL Injection attack that can be launch onto the web application.

Title: Massive distributed and parallel log analysis for organizational security
Authors: X. Shu, J. Smiy, D. Yao, & H. Lin
Description: As network traffic grows and attacks become more prevalent and complex, we must find creative new ways to enhance intrusion detection systems (IDSes). Recently, researchers have begun to harness both machine learning and cloud computing technology to better identify threats and speed up computation times. This paper explores current research at the intersection of these two fields by examining cloud-based network intrusion detection approaches that utilize machine learning algorithms (MLAs). Specifically, we consider clustering and classification MLAs, their applicability to modern intrusion detection, and feature selection algorithms, in order to underline prominent implementations from recent research. We offer a current overview of this growing body of research, highlighting successes, challenges, and future directions for MLA-usage in cloud-based network intrusion detection approaches.

Title: A Novel Unsupervised Anomaly Detection Approach for Intrusion Detection System.
Authors: Weiwei Chen, Fangang Kong
Description: Network Anomaly Detection plays an important part in network security. Among the state-of-the-art approaches, unsupervised anomaly detection is effective when dealing with unlabelled data. However, these approaches also suffer from high false positive rate. We observed that different methods have their own defects and advantages. Inspired by this observation, we provide a new ensemble clustering (NEC) method to detect novel anomalies. In our system, we can get higher detection rate and lower false positive rate compared with existed approaches as verified over NSL-KDD 2009 dataset.

III. PROPOSE SYSTEM

1. The Proposed system will be able to detect the occurred attacks on a particular set of applications.
2. It is a web application for detecting attacks against net Banking based application on input requests by using some heuristic algorithms such as request length, input validation with DFA and anomaly detection for input URL.
3. Each algorithm try to detect certain kind of attacks and total system attacks will be available due to maximum attack detection of each module.
4. Types of attack that are detected in this applications are:

- **SQL injection:** Attackers can use SQL Injection vulnerabilities to bypass application security measures. They can go around authentication and authorization of a web page or web application and retrieve the content of the entire SQL database. They can also use SQL Injection to add, modify, and delete records in the database.
- **DDOS attack:** Victims of a DDoS attack consist of both the end targeted system and all systems maliciously used and controlled by the hacker in the distributed attack.
- **Brute force attack:** In the world of Cyber crimes, brute force attack is an activity which involves repetitive successive attempts of trying various password combinations to break into any website
- **Cross-site attack:** Cross-site Scripting (XSS) is a client-side code injection attack. The attacker aims to execute malicious scripts in a web browser of the victim by including malicious code in a legitimate web page or web application.
- **URL injection:** URL injection is when a malicious individual attacks your website through the insertion of dangerous code that makes it appear as though your website gives credit to a detrimental site.

Attackers can use

The main modules in this system are:-

User:- User has a net banking application. So using application user can register the application .And during registration with the system user have to detail all details correctly that include account detail, personal detail and other necessary things. After registration is successful then login application. Log file save in a cloud. In log file the user information is save.
Admin:-Admin has certain responsibility such as checking user list, search option, Approval part and has IP address.
Log File:- Log file contain Log history, Log URL, Log time, Log date, Log network.
Database:- In the database it will save the information of Log , user data, URL data and history Log.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 8, Issue 4, April 2019

System:- System check the user information , getting network id and LOG URL list. When user login the system and perform the different operation intruder tries to attack the system using various attack. System detect the types of attack that intruder has done and prevent the system for being damaged or data lost.

IV. SYSTEM ARCHITECTURE

Following diagram is our system's architecture

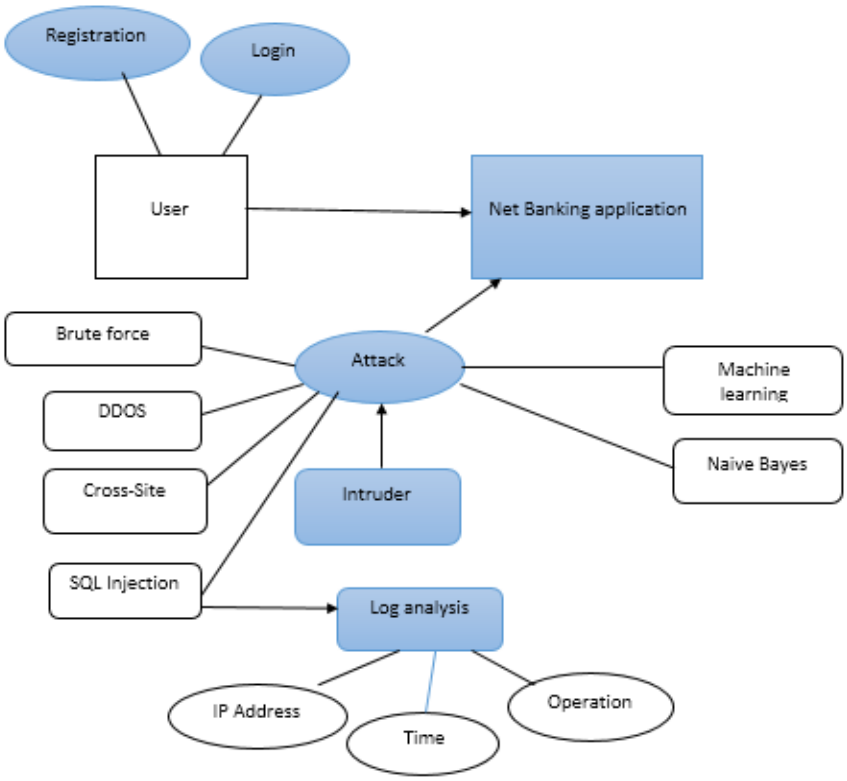

Fig 1: System architecture

V. METHODOLOGY

Mathematical model:
Let _S be the | universal final set
S = {I, O, F, DD, NDD, success, failure}
Identify the inputs as I
I = {I1, I2}
I1= Files
I2 = Receiver's Information
Identify the outputs as O
O = Document access to the receiver
Identify the functions as - F
F = {F1, F2, F3, F4, F5,}
F1 = User registered to the system
F2 = User Login to the system
F3= User perform operation
F4= Attacker tries to damage or attack system F5 = System detect the type of attacks and prevent the system.
Deterministic data = same output from a given starting condition and return the same result any time
Non deterministic data = result will vary every time for given input
Success condition = According to proper inputs
Failure condition = Wrong inputs and ambiguity

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 8, Issue 4, April 2019

Algorithm:-

Naïve Bayes: A naive Bayes classifier algorithm is an algorithm that uses Bayes' theorem to classify objects. Naive Bayes classifiers assume the strong, or naive, independence relation between attributes of data points. Popular uses of naive Bayes classifiers is spam filters, text analysis and medical diagnosis. These classifiers are widely used for machine learning because they are simple to implement. Naive Bayes is also known as simple Bayes or independence Bayes .In this system naïve algorithm is use to detect the different types of attack done on system. So that system can be prevent from the effect of attack .It classify the attack from types of attack and prevent it to harm the system and disturb the user.

VI. CONCLUSION

To With numerous data flowing through the web every day, i-t remains important to detect SQL injection attacks that cause severe security problems on any web-based application hosted on the Internet or on the cloud. A multi-stage log analysis architecture has been proposed, which uses both machine learning and pattern matching methods. Experiment results have shown that the two-stage system is able to detect significantly more SQL injections that a single-stage system. When Bayes Net model (a supervised machine learning method) precedes pattern matching system, the combined system has achieved the best result. Since provides the final output with visualization, it is easy for analysts to further understand, interpret, and take further actions. Future work may include further improvement on queries, using other machine learning methods to reduce manual log classification efforts, and using unsupervised learning methods which may lead to a real-time multi-stage log analysis system Language.

REFERENCES

[1]. David D. Lewis, Feature Selection and Feature Extract ion for Text Categorization, HLT '91 Proceedings of the workshop on Speech and Natural Language Pages 212-217, Chicago
[2]. V. L. Cao, V. T. Hoang, & Q. U. Nguyen. A scheme for building a dataset for intrusion detection systems. In the 2013 Third World Congress on Information and Communication Technologies, pages 120–132, Hanoi- Vietnam, 2013. IEEE
[3]. E. Nasi, "Bypass Antivirus Dynamic Analysis: Limitations of the AV model and how to exploit them," Aug. 2014.
[4]. K. Boudaoud, Detection of intrusions: A new approach by multi - agents systems, thesis of doctorate,Lausanne, EPFL 2002.
[5]. MEI-LING SHYU, THIAGO QUIRINO, and ZONGXING XIE. (2007). Network Intrusion Detection through Adaptive Sub-Eigenspace Modeling in Multiagent Systems. ACMTransactions on Autonomous and Adaptive Systems. 2 (n.d), p-1-37.
[6]. A. Patwardhan , J. Parker , M. Iorga , A. Joshi , T.Karygiannis and Y. Yesha. (2008). Threshold-based intrusion detection in ad hoc networks and secureAODV. elsevier. 6 (n.d), p-578–599.
[7]. S.A. Razak , S.M. Furnell , N.L. Clarke and P.J. Brooke. (2008). Friend-assisted intrusion detection and response mechanisms for mobile ad hoc networks. elsevier. 6 (n.d), p-1151–1167.
[8]. Ningrinla Marchang and Raja Datta. (2008). Collaborative techniques for intrusion detection in mobile ad-hoc networks. elsevier. 6 (n.d), p-508-523.
[9]. Hadi Otrok , Noman Mohammed, Lingyu Wang, Mourad Debbabi and Prabir Bhattacharya. (2008). A game-theoretic intrusion detection model for mobile adhoc networks. elsevier. 31 (n.d), p-708–721.