Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Detecting Web Attacks Using Deep Learning

Dr. B. Sanjai Prasada Rao

Associate Professor

Department of Computer Science and Engineering MLR Institute of Technology Hyderabad, India-500043

Sathwik Annaldas

Department of CSE MLR Institute of Technology Hyderabad, India-500043

Meghana Arukala

Department of CSE MLR Institute of Technology Hyderabad, India-500043

A. Rishwanth

Department of CSE MLR Institute of Technology Hyderabad, India-500043

Article Info
Page Number: 11168-11174
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received:15 September 2022
Revised: 25 October 2022
Accepted: 14 November 2022

Abstract: Web attack is a cyber-attack which is used to steal the sensitive information and Un authorization of the computer system without knowing oneself. Most of the cyber attackers' targets web applications because they can be accessed easily. When attack is happened an intrusion detection system sends the alert message to the users and detected. The implementation of existing methods can be time consuming and more expensive and requires a lot of domain knowledge. By our prevention method we use RSMT model to monitor the runtime behaviour of web application. Then, RSM tool trains an auto encoder which is evoked to encode the unlabelled data which is used to identify anomalies. After these results of both datasets application is analysed. Our proposed system can detect the cyber-attack with small amount of labelled training data

Introduction

Web applications are the major sort of target for cyber attackers. The attackers mostly use the SQL injection, cross site scripting. These attacks can stop the web services and can steal the sensitive information from users which will lead to the major economic loss to users as well as service providers. Even through there are lot of securing services which were developed by developers are impossible to protect some web applications The methods like conventional intrusion detection system does not work as expected due to these below mentioned reasons .For the detection of cyber-attack an security expert is required to identify the features that are related to draw out from the network packages. Many cyber-attack detection systems depend on the supervised machine learning algorithms to identify the typical attacks. This required the large amount of labelled trained data to train algorithms which is expensive and time consuming. Besides supervised learning practices can identify the existing attacks. But the cyber attackers always find the new form for the intrusion of data. This may lead to mischaracterization. The unsupervised learning algorithms (SVM) is also used to detect the web attacks, but this method requires the manual selection of typical attacks.

Iii. Literature Survey

A classification of SQL-injection attacks and countermeasures

AUTHORS: Halfond WG, Viegas J, Orso A.

ABSTRACT:SQL injection attacks pose a serious security threat to Web applications: they allow attackers to obtain unrestricted access to the databases underlying the applications and to the potentially sensitive information these databases contain. Although researchers and practitioners have proposed various methods to address the SQL injection problem, current approaches either fail to address the full scope of the problem or have limitations that prevent their use and adoption.

An adaptive network intrusion detection method based on pca and support vector machines. Advanced Data Mining and Applications.

AUTHORS: Xu X, Wang X.

ABSTRACT: Network intrusion detection is an important technique in computer security. However, the performance of existing intrusion detection systems (IDSs) is unsatisfactory since new attacks are constantly developed, and the speed of network traffic volumes increases fast. To improve the performance of IDSs both in accuracy and speed, this paper proposes a novel adaptive intrusion detection method based on principal component analysis (PCA) and support vector machines (SVMs).

Using Adaptive Alert Classification to Reduce False Positives in Intrusion Detection

AUTHORS: Pietraszek T.

ABSTRACT: Intrusion Detection Systems (IDSs) are used to monitor computer systems for signs of security violations. Having detected such signs, IDSs trigger alerts to report them. These alerts are presented to a human analyst, who evaluates them and initiates an adequate response. In practice, IDSs have been observed to trigger thousands of alerts per day, most of which are false positives (i.e., alerts mistakenly triggered by benign events).

An anomaly detection method to detect web attacks using stacked auto-encoder

AUTHORS: Ali Moradi Vartouni,Saeed Sedighian Kashi, Mohammad Technilab

ABSTRACT: Network borne attacks are currently major threats to information security. Enormous efforts such as scanners, encryption devices, intrusion detection systems and firewalls have been made to mitigate these attacks. Web application firewalls use intrusion detection techniques to protect servers form HTTP traffic and, Machine learning algorithms have used based on anomaly detection in these firewalls.

IV. LIMITATIONS OF EXISTING SYSTEMS

Lack of identifying attacks in web application.
Researchers found that more than half of web applications during a 2015-2016 scan contained high security vulnerabilities

V.PROPOSED SYSTEM

The proposed system first, we evaluate the datasets of an unsupervised and semi-supervised approach towards web attack detection based on the Robust Software Modelling Tool.

RSMT is a tool which can efficiently look over the behaviour of system which will record all the executing trace of datasets and trace files. Trace file contains raw data which is low dimensional it cannot be used for Deep learning Network. To make it useful for Deep Learning Models we use Auto-encoder technique. Auto encoder technique will convert low dimensional data into deep learning features.

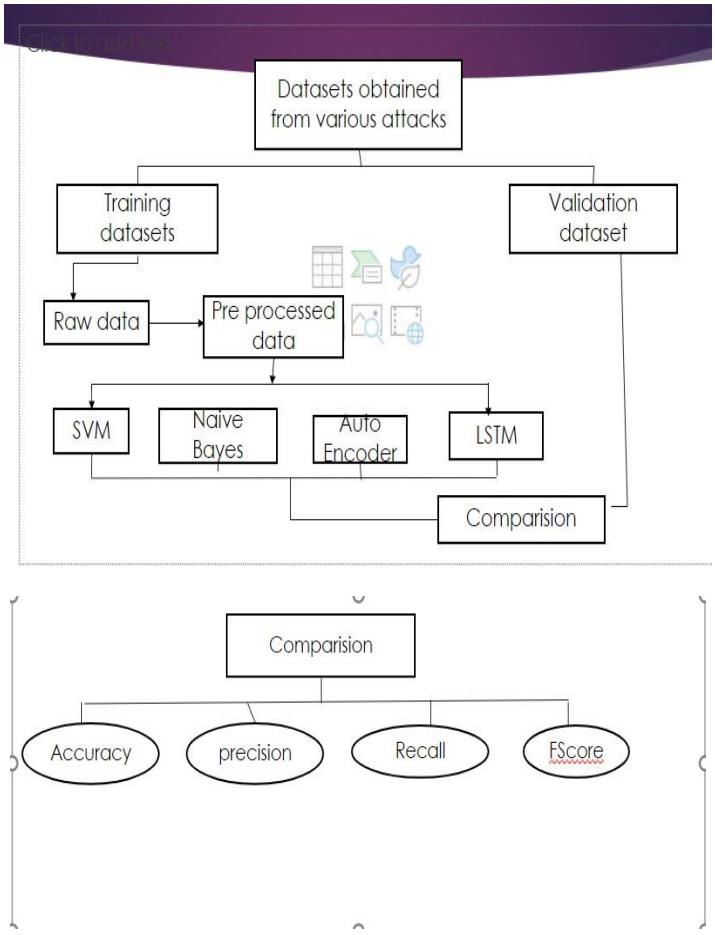
These features will be passes to propose Auto-Encoder algorithm which will generate two models

Train datasets

Test datasets

New, Test datasets will be applied to Trained model to identify weather new test data is a normal request or contains any attacks. If no test data is available in the auto-encoder trained model, we consider it as attack.

PROPOSED SYSTEM ARCHITECTURE



We calculate mainly four features.

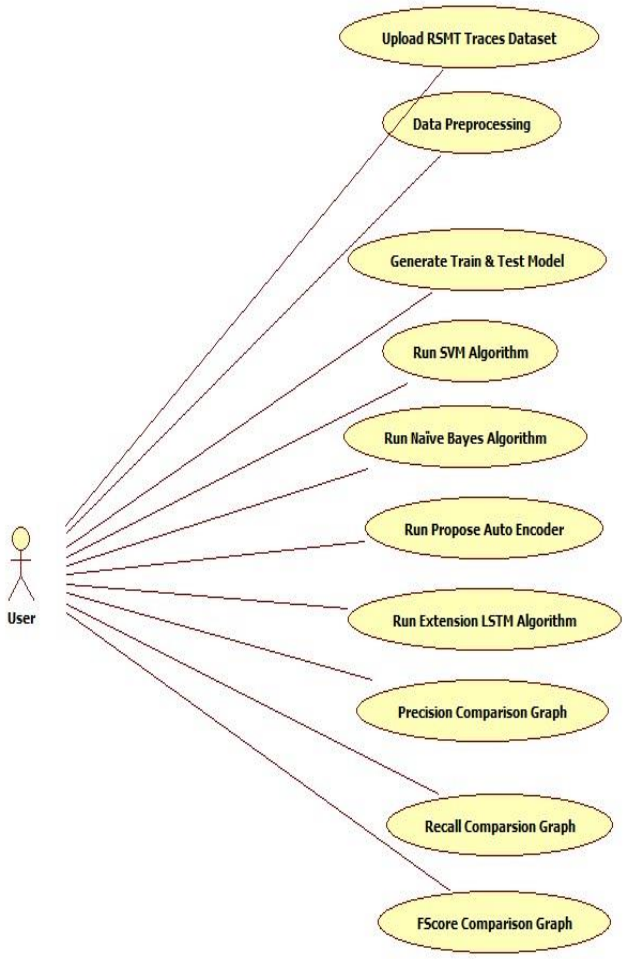
1 Accuracy

2 Precision

3 FS core
4 Recall.
Using LSTM, Auto Encoder Algorithms we construct Call Graphs for Accuracy, Precision, Recall and FS core, where low dimensional call graphs are considered as attacks and high dimensional call graphs value will be higher or closer to 100 is considered as Safe.

B. UML Diagram

The UML diagram shows the interaction of one component with another in a linear fashion .



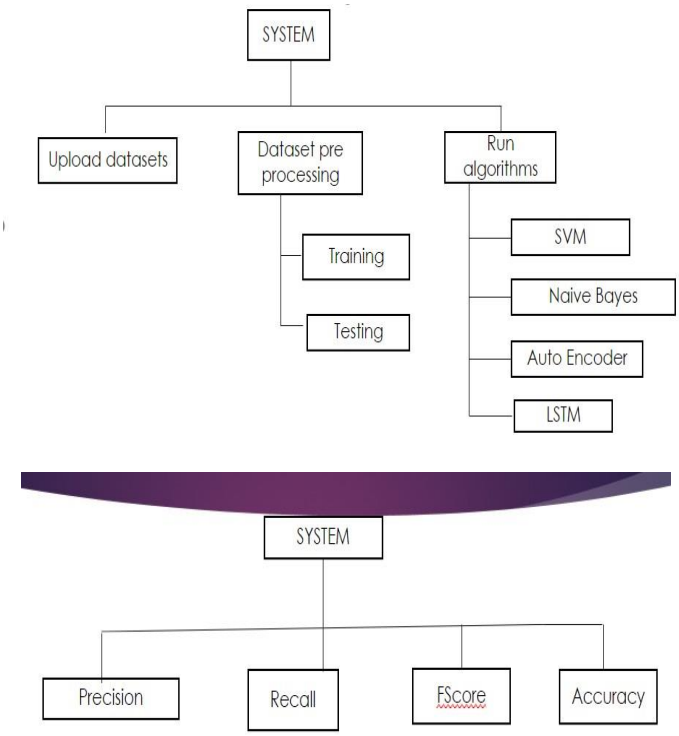
C. Module Diagram

The module diagram shows a high-level view on the flow of execution of project. It has been categorised into 3 phases:

Upload Datasets

Dataset Preprocessing

Run Algorithms



Upload datasets: using this module we will upload the datasets.

Dataset pre-processing: using this module we will find out empty values in the dataset and replace with mean values or 0 values.

Train and Test Split: Using this module we will split dataset into two parts called training and test data. 80% dataset to train classifier and 20% dataset is used to test classifier.

Run SVM algorithm: Using this module we will train SVR classifier with splitted 80% data and used 20% data to calculate it performance.

Run Naive Bayes algorithm: Using this module we will train Naive Bayes classifier with splitted 80% data and used 20% data to calculate it performance.

Vi.Simulation Settings
A. Hardware Requirements

System : Pentium IV 2.4 GHz

Hard Disk : 40 GB

Floppy Drive : 1.44 Mb

Monitor : 15 VGA Colour

Mouse : Logitech

Ram : 512 Mb

Software Requirements

Operating System: Windows

Coding Language: Python 3.

VII. Conclusion

This project explains the result of applying unsupervised learning datasets for a deep learning that can detect the attacks on the web application automatically. We have built the web application using RSMT tool, which defines the runtime behaviour of web applications and monitors it. We will apply an auto encoder learning reduced dimensions rendition of extracted data from runtime application. To analyse the function of our intrusion detection system, we have designed Test applications and then the efficiency can be evaluated of unsupervised learning averse to datasets

Using LSTM, Auto Encoder Algorithms we construct Call Graphs for Accuracy, Precision, Recall and FScore, where low dimensional call graphs are considered as attacks and high dimensional call graphs value will be higher or closer to 100 is considered as Safe.

Acknowledgment

It would have been highly difficult for such a project to be carried out without the help of our guide Dr. B. Sanjai Prasada Rao, his guidance and help made the project to be driven in the correct path with minimal scope of error. I would also like to thank all the authors of the IEEE papers we have referred to for doing this project, their results helped us to select the correct choices and actions.

References

1. Halfond WG, Viegas J, Orso A. A classification of sql-injection attacks and countermeasures. In: Proceedings of the IEEE International Symposium on Secure Software Engineering. IEEE: 2006. p. 13–5.
2. Wassermann G, Su Z. Static detection of cross-site scripting vulnerabilities. In: Proceedings of the 30th International Conference on Software Engineering. ACM: 2008. p. 171–80.
3. Di Pietro R, Mancini LV. Intrusion Detection Systems vol. 38: Springer; 2008.
4. Qie X, Pang R, Peterson L. Defensive programming: Using an annotation toolkit to build dos-resistant software. ACM SIGOPS Oper Syst Rev. 2002; 36(SI):45–60.
5. https://doi.org/https://www.acunetix.com/acunetix-web-application-vulnerability-report-2016. Accessed 16 Aug 2017.
6. https://doi.org/http://money.cnn.com/2015/10/08/technology/cybercrime-cost-business/index.html. Accessed 16 Aug 2017.
7. https://doi.org/https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do. Accessed 16-August-2017.
8. https://doi.org/https://theconversation.com/why-dont-big-companies-keep-their-computer-systems-up-to-date-84250. Accessed 16 Aug 2017.

9. Ben-Asher N, Gonzalez C. Effects of cyber security knowledge on attack detection. Comput Hum Behav. 2015; 48:51–61.
10. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intell Data Anal. 2002; 6(5):429–49
11. Liu G, Yi Z, Yang S. A hierarchical intrusion detection model based on the pca neural networks. Neurocomputing. 2007; 70(7):1561–8.
12. Xu X, Wang X. An adaptive network intrusion detection method based on pca and support vector machines. Advanced Data Mining and Applications. 2005; 3584:696–703.
13. Pietraszek T. Using adaptive alert classification to reduce false positives in intrusion detection. In: Recent Advances in Intrusion Detection. Springer: 2004. p. 102–24.
14. Goodfellow I, Bengio Y, Courville A. Deep Learning: MIT press; 2016.
15. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.: 2012. p. 1097–105.