ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Detecting Web Attacks with end-to-end Deep Learning

P. Bhaskar, A. Venkata Subbaiah, J. Prathap,M.Benhur Harrison,S. Jnaneswar, B.Md. Sohail, G.Vijaya Rahul
Dept of C.S.E,SREC, Nandyal.
bhaskar.cse@srecnandyal.edu.in ,subbaiah.cse@srecnandyal.edu.in,

Abstract- Web applications are popular targets for cyber-attacks because they are network-accessible and often contain vulnerabilities. An intrusion detection system monitors web applications and issues alerts when an attack attempt is detected. Existing implementations of intrusion detection systems usually extract features from network packets or string characteristics of input that are manually selected as relevant to attack analysis. Manually selecting features, however, is time-consuming and requires in-depth security domain knowledge. Moreover, large amounts of labeled legitimate and attack request data are needed by supervised learning algorithms to classify normal and abnormal behaviors, which is often expensive and impractical to obtain for production web applications.This project provides three contributions to the study of autonomic intrusion detection systems. First, we evaluate the feasibilityof an unsupervised/semi-supervised approach for web attack detection based on the Robust Software Modeling Tool (RSMT), which autonomically monitors and characterizes the runtime behavior of web applications. Second, we describe how RSMT trains a stacked denoising autoencoder to encode and reconstruct the call graph for end-to-end deep learning, where a low-dimensional representation of the raw features with unlabeled request data is used to recognize anomalies by computing the reconstruction error of the request data. Third, we analyze the results of empirically testing RSMT on both synthetic datasets and production applications with intentional vulnerabilities. Our results show that the proposed approach can efficiently and accurately detect attacks, including SQL injection, cross-site scripting, and deserialization, with minimal domain knowledge and little labeled training data. In this project author evaluating propose Auto Encoder Algorithm with SVM and Naïve Bayes. In extension work we are using LSTM algorithm and comparing with all algorithms.

Keywords- RSMT,LSTM etc.

I. INTRODUCTION

Emerging trends and challenges. Web applications are attractive targets for cyber attackers. SQL execution are common attacks that can disable web services, steal sensitive user information, and cause significant financial loss to both service providers and users. Protecting web applications from attack is hard. In this developed system, the files are tested with different deep learning algorithms to findthe best algorithm that produce more accuracy in detecting web attacks. This is done by testing the files with all the algorithms . Here, we already know the number of files got attacked before testing them with algorithms. Now , by looking at the accuracies of each algorithm we can find the best algorithm for detecting web attacks.

II.METHODOLOGY

- Upload RSMT Traces Dataset is the first module of the our project and it is used to upload attack traces.
- Generate Train & Test Model module is used to generate train and test data. All deep learning algorithms will take 80% dataset as training and 20% dataset for testing.
- Run SVM Algorithm module is used to generate SVM model on train data and calculate precision, recall on test data.
- Run Naive Bayes Algorithm module is used to generate Naive Bayes model on train data and calculate precision, recall on test data.
- Run Propose Auto Encoder module to run propose algorithm, Auto Encoder got 90% accuracy.
- Run Extension LSTM Algorithm module is used to run LSTM and we got values for LSTM algorithm.
- Precision Comparison graph x-axis represents algorithm name and y-axis represents precision value. In all algorithm propose Auto Encoder showing good performance.
- Recall Comparison graph x-axis represents algorithm name and y-axis represents recall value. In all algorithm Extension LSTM showing good performance.
- FScore comparison graph x-axis represents algorithm name and y-axis represents FScore value. In all algorithm Extension LSTM showing good performance.

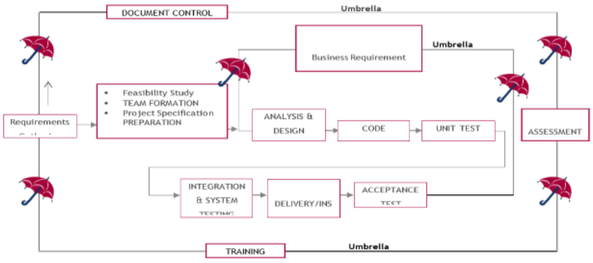


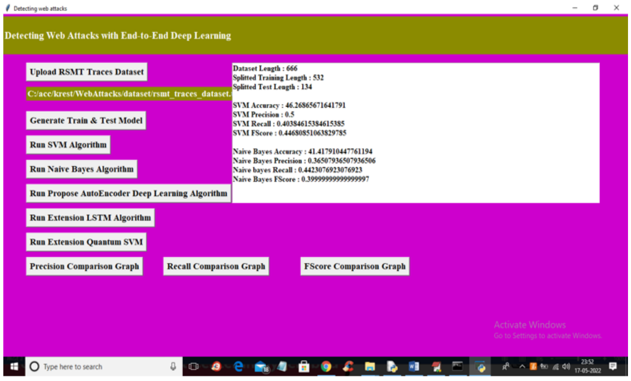
Fig. 1 SDLC (Umbrella Model)

1.Processing System

Upload RSMT Traces Dataset is the first step of the our project and it is used to upload attack traces. Generate Train & Test Model module is used to generate train and test data. All deep learning algorithms will take 80% dataset as training and 20% dataset for testing. We ran SVM Algorithm that is used to generate SVM model on train data and calculate precision, recall on test data and next we ran Naive Bayes Algorithm that is used to generate Naive Bayes model on train data and calculate precision, recall on test data. Ran Propose Auto Encoder Algorithm on the dataset, Auto Encoder got 90% accuracy. Next, we made to run Extension LSTM Algorithm that is used to run LSTM and we got values for LSTM algorithm.Precision Comparison graph x-axis represents algorithm name and y-axis representsprecision value. In all algorithms propose Auto Encoder showing good performance and FScore comparison graph x-axis represents algorithm name and y-axis represents FScore value. In all algorithm Extension LSTM showing good performance.

III.RESULT AND DISCUSSION

Upload the dataset by tapping on the 'Upload RSMT data' button. Train and test data by clicking 'Generate Train and Test' button. Now, we have to run all the alogrithms on the dataset.

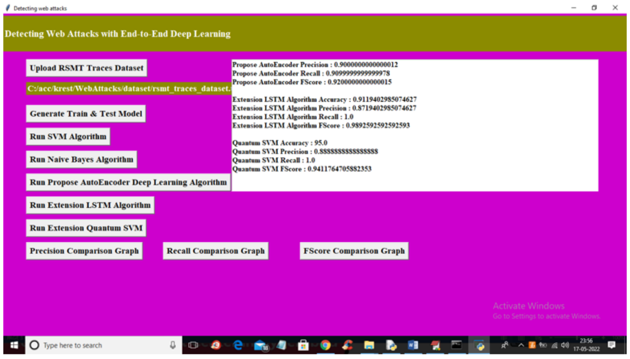

Detecting web attacks
Detecting Web Attacks with End-to-End Deep Learning
Upload RSMT Traces Dataset
C:/acc/krest/WebAttacks/dataset/rsmt_traces_dataset.
Generate Train & Test Model
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Propose AutoEncoder Deep Learning Algorithm
Run Extension LSTM Algorithm
Run Extension Quantum SVM
Precision Comparison Graph
Recall Comparison Graph
FScore Comparison Graph
Activate Windows
Go to Settings to activate Windows.
Type here to search
23:52
17-05-2022

```
Dataset Length : 666
Splitted Training Length : 532
Splitted Test Length : 134

SVM Accuracy : 46.26865671641791
SVM Precision : 0.5
SVM Recall : 0.40384615384615385
SVM FScore : 0.44680851063829785

Naive Bayes Accuracy : 41.417910447761194
Naive Bayes Precision : 0.36507936507936506
Naive bayes Recall : 0.4423076923076923
Naive Bayes FScore : 0.39999999999999997
```

Fig.2 Upload RSMT Data


Detecting web attacks
Detecting Web Attacks with End-to-End Deep Learning
Upload RSMT Traces Dataset
C:/acc/krest/WebAttacks/dataset/rsmt_traces_dataset.
Generate Train & Test Model
Run SVM Algorithm
Run Naive Bayes Algorithm
Run Propose AutoEncoder Deep Learning Algorithm
Run Extension LSTM Algorithm
Run Extension Quantum SVM
Precision Comparison Graph
Recall Comparison Graph
FScore Comparison Graph
Activate Windows
Go to Settings to activate Windows.
Type here to search
23:56
17-05-2022

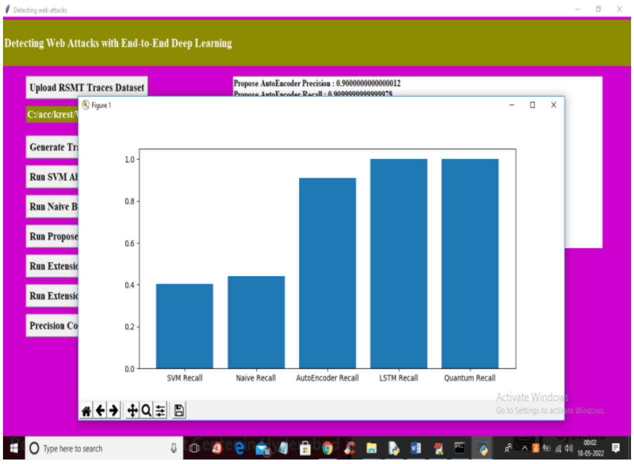
```
Propose AutoEncoder Precision : 0.9000000000000012
Propose AutoEncoder Recall : 0.9099999999999978
Propose AutoEncoder FScore : 0.9200000000000015

Extension LSTM Algorithm Accuracy : 0.9119402985074627
Extension LSTM Algorithm Precision : 0.8719402985074627
Extension LSTM Algorithm Recall : 1.0
Extension LSTM Algorithm FScore : 0.9892592592592593

Quantum SVM Accuracy : 95.0
Quantum SVM Precision : 0.8888888888888888
Quantum SVM Recall : 1.0
Quantum SVM FScore : 0.9411764705882353
```

Fig. 3 Recall Graph.

In the above pictures we can see the accuracies of the algorithms that are used in the project. We have also used graphs for comparing Precison. Recall, F Score of the results from each algorithm.



```
Propose AutoEncoder Precision : 0.9000000000000012
```

Fig. 4 Alogrithms

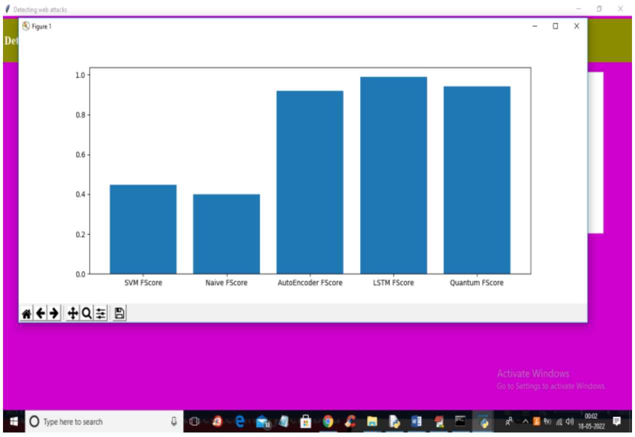


Fig.5 FSCORE graph.

From the accuracies that have given by each algorithm .We can find out the algorithm that gives the best output on detecting web attacks.

IV.CONCLUSION

Thus, we have developed a system that produces best algorithm in detecting web attacks and it's clear that Quantum SVM algorithm is having the highest accuracycomparing to others. So, we can conclude that with Quantum SVM algorithm we can detect web attacks with more accuracy than any other algorithms.

REFERENCES

[1] MV Subramanyam, K Soundararajan, J. Sofia Priya Dharshini"Adaptive Modulation and Coding With Incremental Redundancy Hybrid ARQ in MIMO Systems: A Cross Layered Design.", International Journal of Engineering Research and Applications, Vol.3, no.5, pages. 503-7,2013.

[2] MV Subramanyam, K Satyaprasad, S L Prathapa Reddy"A hybrid genetic fuzzy approach for power control cross layer MAC protocol in wirelessnetwork", International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) , pages, 181-186, December2015.

[3] MV Subramanyam ,R Sumalatha"Image denoising usingSpatialAdaptiveMaskFilter for medical images", International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), pages. 1-4, June2015.

[4] MakamVenkataSubramanyam,KodatiSatyaPrasad, BandaniAnilKumar"Anenergy efficient clustering using K-Means and AODV routing protocol in Ad-hoc networks" , Vol.12, no.2, pages. 125-134,2019.

[5] Farooq Sunar Mahammad, M. Sharmila Devi, D Bhavana, D Sukanya, TV Sai Thanusha, M Chandrakala, P Venkata Swathi "Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection" JOURNAL OF ALGEBRAIC STATISTICS, Vol.13, no.3, pages. 112-117, June2022.

[6] SunarmohammedFarooq,"StaticPeersforPeer-to-PeerLiveVideoStreaming'',Inter national journal of Scientific Engineering and Technology Research, Vol.05, No.34, Pages:7055-7064,October-2016.

[7] Farooq Sunar Mahammad,Palanisamy Ramasamy, Karthik Balasubramanian "Comparative analysis of 3D-SVM and

4D-SVM for five-phase voltage sourceinverter”, International Transactions on Electrical Energy Systems, Vol.31, No.12, Pages: e13138, December-2012.

[8] P Bhaskar, Farooq Sunar Mahammad, A Hemanth Kumar “Machine Learning Based PredictiveModelforClosedLoopAirFilteringSystem”,JOURNALOFALGEBRAIC STATISTICS, Vol.13, no.3, pages. 609-616, July2022.

[9] P Bhaskar, A Prudvi ,N Yugandhar Reddy “PredictionOfCovid-19 Infection Based on Life style Habits Employing Random Forest Algorithm” ,JOURNAL OF ALGEBRAIC STATISTICS, Vol.13, no.3, pages. 40-45, June2022.

[10] M. Amareswara Kumar, FarooqSunar Mahammad“Traffic Length Data Based Signal Timing Calculation for Road Traffic Signals Employing Proportionality Machine Learning”JOURNAL OF ALGEBRAIC STATISTICS, Vol.13, no.3, pages. 40-45, June 2022.

[11] K.V. Sai Phani, RAJESH SATURI “A FRAME WORK TO DETECTBREAST CANCER USING KNN and SVM” European Journal of Molecular & Clinical Medicine, Vol.8, no.3, pages. 1432-1438, 2021.

[12] V Lakshmi Chaitanya, “Machine Learning Based Predictive Model for Data Fusion BasedIntruder Alert System”,journal of algebraic statistics,Vol.13,no.2,pages.2477- 2483, June2022.

[13] 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.

[14] Sun F, Zhang P, White J, Schmidt D, Staples J, Krause L. A feasibility study of autonomically detecting in-process cyber-attacks. In: Cybernetics (CYBCON), 2017 3rd IEEE International Conference On. IEEE; 2017. p. 1–8.

[15] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res. 2010;11(Dec): 3371–408.

[16] Fu X, Lu X, Peltsverger B, Chen S, Qian K, Tao L. A static analysis framework for detecting sql injection vulnerabilities. In: Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International. IEEE; 2007. p. 87–96.