IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 5, May 2024

Semisupervised Machine Learning Approach for Ddos Detection

Sanjeevi. J[1] and Dr. Krithika. D. R.[2]

PG Student, Department of Computer Applications[1]

Assistant Professor, Department of Computer Applications[2]

Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India

22304139@vistas.ac.in and krithikabanu@gmail.com

Abstract*: Analyzing cyber incident data sets is an important method for deepening our understanding of the evolution of the threat situation. In present generation we come to know about many cyber breaches and hacking taking place. In this project work, we research about the various cyber- attacks and breaches and study the way these attacks are done and find an alternative for the same. We show that rather than by distributing these attacks as because they exhibit autocorrelations, we should model by stochastic process both the hacking breach incident inter- arrival times and breach sizes. We draw a set of cyber securities insights, including that the threat of cyber hacks is indeed getting worse in terms of their frequency. In our project we will be using the algorithms such as Convolution Neural Network (CNN) as existing and Recurrent Neural Network (RNN) as proposed for analyzing our results. From the results obtained its proved that proposed RNN works better than existing CNN.*

Keywords: DDoS (Distributed Denial of Service) Deduction, Machine Learning, Semi-Supervised Learning, Network Security

I. INTRODUCTION

An information rupture is a security occurrence in which delicate, ensured or secret information is duplicated, transmitted, saw, stolen or utilized by an individual unapproved to do as such." An information break is the purposeful or accidental arrival of secure or private/classified data to an untrusted domain. Different expressions for this marvel incorporate inadvertent data divulgence, information spill and furthermore information spill. This may incorporate occurrences, for example, robbery or loss of advanced media, for example, PC tapes, hard drives, or smart phones such media whereupon such data is put away decoded, posting such data on the internet or on a PC generally available from the Internet without legitimate data security safeguards, exchange of such data to a framework which isn't totally open yet isn't fittingly or formally authorize for security at the affirmed dimension, for example, decoded email - or exchange of such data to the data frameworks of a conceivably unfriendly office, for example, a contending organization or a remote country, where it might be presented to increasingly serious unscrambling strategies. While mechanical arrangements can solidify digital frameworks against assaults, information breaks keep on being a major issue.

This propels us to describe the development of information rupture occurrences. This not exclusively will profound our comprehension of information breaks, yet in addition shed light on different methodologies for relieving the harm, for example, protection. Many trust that protection will be valuable, however the advancement of accurate cyber hazard measurements to control the task of protection rates is past the compass of the present comprehension of information breaks In this paper, we make the accompanying commitments. We shoe that as opposed to by circulating the ruptures we should demonstrate by stochastic procedure both the hacking break occurrence entomb entry times and rupture sizes.

II. LITERATURE SURVEY

Detecting Port Scan Attempts with Comparative Analysis of Deep Learning and Support Vector Machine Algorithms Dogukan Aksu , M. Ali Aydin IEEE 2023 [1]. Detecting cyber-attacks using a CRPS-based monitoring approach Fouzi Harrou ; Benamar Bouyeddou ; Ying Sun ; Benamar Kadri IEEE 2023. [2]. A Taxonomy of Malicious Traffic for

DOI: 10.48175/IJARSCT-18409

Intrusion Detection Systems Hanan Hindy ; Elike Hodo ; Ethan Bayne ; Amar Seeam ; Robert Atkinson ; Xavier Bellekens IEEE 2023. [3]. Parameter-Invariant Monitor Design for Cyber–Physical Systems James Weimer ; RadoslavIvanov ; Sanjian Chen ; Alexander Roederer ; Oleg Sokolsky ; Insup Lee IEEE 2023. [4]. A novel approach for terrorist sub-communities detection based on constrained evidential clustering Firas Saidi ; Zouheir Trabelsi ; Henda Ben Ghazela IEEE 2023. [5]. Causality Countermeasures for Anomaly Detection in Cyber-Physical Systems Dawei Shi ; Ziyang Guo ; Karl Henrik Johansson ; Ling Shi IEEE 2023. [6]. Proactive cyber security response by utilizing passive monitoring technologies Koji Nakao IEEE 2023. [7 Simplistic Approach to Detect Cybercrimes and Deter Cyber Criminals Abu Shakil Ahmed ; Sudip Deb ; Al-Zadid Sultan Bin Habib ; Md. Nurunnabi Mollah ; Abu Saleh Ahmad IEEE 2023. [8]. Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for Credit Card Fraud Detection Ibtissam Benchaji ; Samira Douzi ; Bouabid ElOuahidi IEEE 2023. [9]. Defending against common cyber attacks: Phishing and cross-site scripting Al-Sakib Khan Pathan IEEE 2023. [10].

III. METHODOLOGY SECTION AND EXPERIMENTAL RESULT

Module Description

Input dataset:

Dataset can be taken from online source provider called UCI repository. We have collected set of criminal datasets which we are going to analyze. Then for training the data set also for the comparison of the non-criminal datasets are also been taken.

Analysis of data set:

Here the analysis if dataset takes place. The size of data is taken into consideration for the data process.

Oversampling (Using SMOTE): we have created a detailed history of all crimes that been complained over a certain amount of time and it is sampled to fix a threshold value.

Training and Testing Subset: As the dataset is imbalanced, many classifiers show bias for majority classes. The features of minority class are treated as noise and are ignored. Hence it is proposed to select a sample dataset.

Applying algorithm: Following are the classification algorithms used to test the sub-sample dataset.

- Convolution Neural Network (CNN) and
- Recurrent Neural Network (RNN)

Predicting results: The test subset is applied on the trained model .The metrices used is accuracy. The ROC Curve is plotted and the desirable results are achieved.

IV. SYSTEM ARCHITECTURE

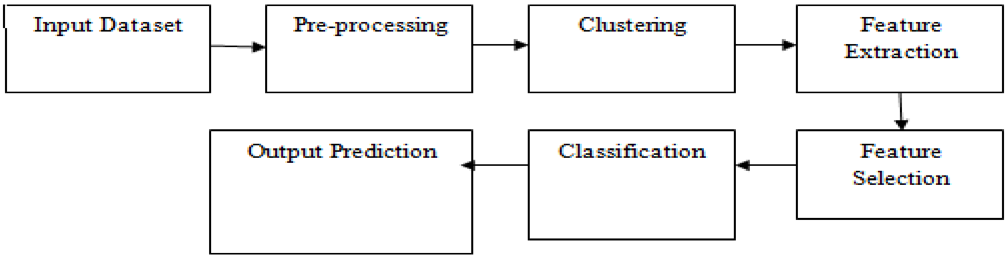


Fig.1 System Architecture

DOI: 10.48175/IJARSCT-18409

V. DATA FLOW DIAGRAM

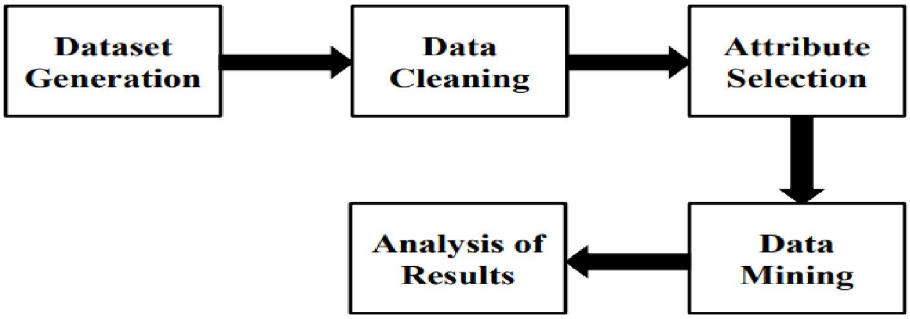
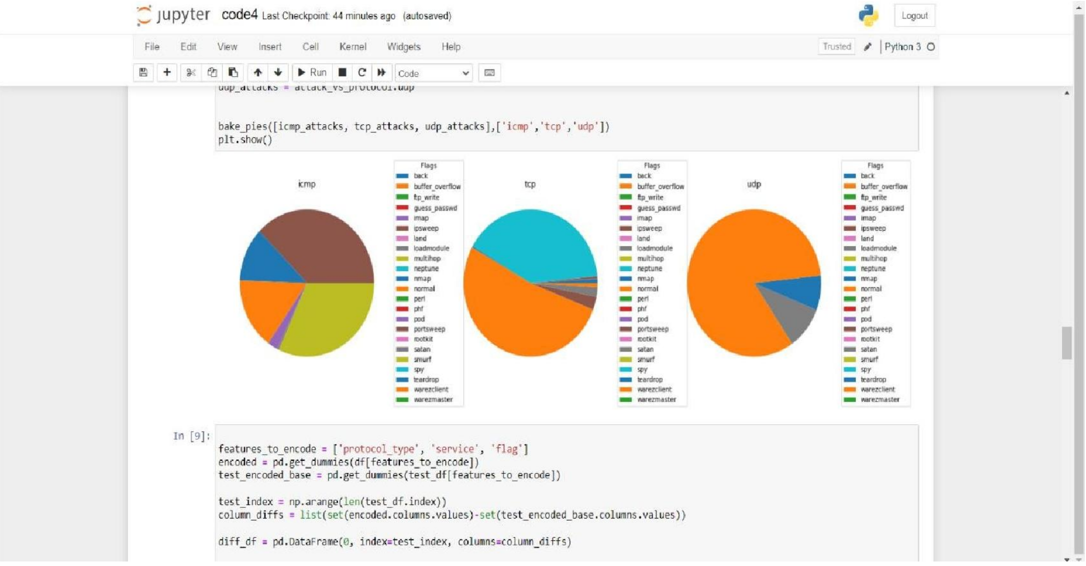


Fig 2. Data Flow Diagram



```
udp_attacks = attack_vs_protocol.udp


bake_pies([icmp_attacks, tcp_attacks, udp_attacks],['icmp','tcp','udp'])
plt.show()
```

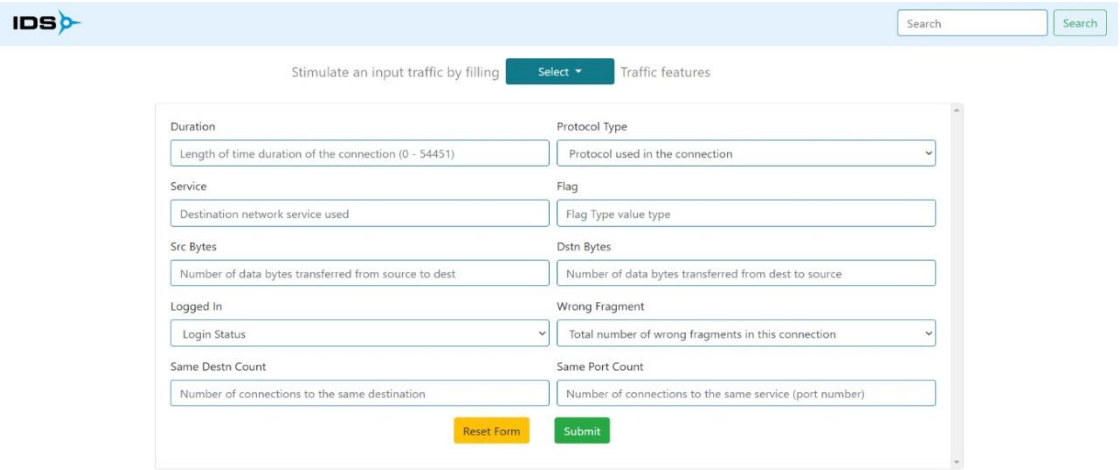
```
In [9]:

features_to_encode = ['protocol_type', 'service', 'flag']
encoded = pd.get_dummies(df[features_to_encode])
test_encoded_base = pd.get_dummies(test_df[features_to_encode])

test_index = np.arange(len(test_df.index))
column_diffs = list(set(encoded.columns.values)-set(test_encoded_base.columns.values))

diff_df = pd.DataFrame(0, index=test_index, columns=column_diffs)
```

Fig 3. Data exploration



Stimulate an input traffic by filling Select Traffic features

Duration: Length of time duration of the connection (0 - 54451)

Protocol Type: Protocol used in the connection

Service: Destination network service used

Flag: Flag Type value type

Src Bytes: Number of data bytes transferred from source to dest

Dstn Bytes: Number of data bytes transferred from dest to source

Logged In: Login Status

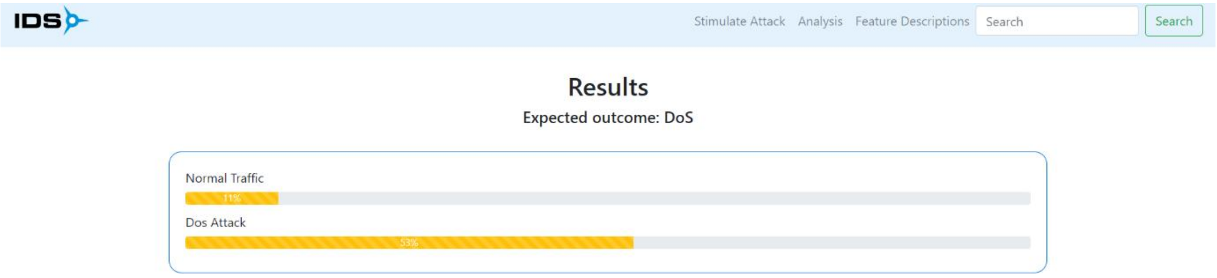
Wrong Fragment: Total number of wrong fragments in this connection

Same Destn Count: Number of connections to the same destination

Same Port Count: Number of connections to the same service (port number)

Fig.4. Inputs

DOI: 10.48175/IJARSCT-18409



Results

Expected outcome: DoS

Normal Traffic 11%

Dos Attack 53%

Fig5. Results

V. CONCLUSION

The widespread of ordinary data breaches around the world demonstrates how real the danger of critical infrastructure attack As the hackers increase in terms of sophistication and technical expertise, and as the critical information infrastructure becomes more massive and intricate, it is more vulnerable to attack. We can treat them like an act of terrorism which justifies action under the Internal Security Act. If we take this path, we must be prepared of the consequences. What is more compelling is the need to strengthen the security itself. As illustrated in this article, a multi-prong action is required; one that involves a mixture of technology, competency of manpower, prudence and effective legal framework. At this end, it is note-worthy that there are few areas emerged from this initial study that can be made an agenda of future direction.

Firstly, from the technical perspective, there is a need to assess new methods that threaten the security of critical information infrastructure. Secondly, from the perspective of law and policy, governments need to ensure that each sector identified as critical infrastructure should be properly protected both by legal and policy instruments.

REFERENCES

[1]. P. R. Clearinghouse. Privacy Rights Clearinghouse's Chronol- ogy of Data Breaches Accessed: Nov. 2018. [Online]. Available: https://www.privacyrights.org/data-breaches

[2]. ITR Center. Data Breaches Increase 40 Percent in 2019, Finds New Report From Identity Theft Resource Center and CyberScout. Accessed: Nov. 2018. [Online]. Available: http://www.idtheftcenter.org/2016databreaches.html

[3]. C. R. Center. Cybersecurity Incidents. Accessed: Nov. 2020. [Online]. Available: https://www.opm.gov/cybersecurity/cybersecurity-incidents

[4]. IBM Security. Accessed: Nov. 2020. [Online]. Available: https://www.ibm.com/security/data-breach/index.html

[5]. NetDiligence. The 2016 Cyber Claims Study. Accessed: Nov. 2019. [Online]. Available:https://netdiligence.com/wpcontent/uploads/2016/10/P02_NetDiligence- 2018.

[6]. Vidushi Sharma ,Sachin Rai, Anurag Dev" A Comprehensive Study of Artificial Neural Networks" International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 10, October 2012.

[7]. Nabil EL KADHI, Karim HADJAR, Nahla EL ZANT " A Mobile Agents and Artificial Neural Networks for Intrusion Detection" Journal Of Software, VOL. 7, NO. 1, JANUARY 2012.

[8]. Linda Ondrej, T. Vollmer, M. Manic, (2009) "Neural Network Based Intrusion Detection System for Critical Infrastructures", Proceedings of International Joint Conference on Neural Networks, pp. 1827 1834.

DOI: 10.48175/IJARSCT-18409