International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 6 Volume No.6 October - November – 2022
DOI:10.46647/ijetms.2022.v06i06.072 ISSN: 2581-4621

NETWORK TRAFFIC DETECTION THROUGH MACHINE LEARNING

Asiya Begum[1], B.Srinivas S.P Kumar[2]
[1] *MCA IV Semester, Dept. of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad – 500 075, India*
[2] *Assistant Professor, Dept. of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad – 500 075, India*

ABSTRACT

In imbalanced network traffic, malicious cyber-attacks can often hide in large amounts of normal data. It exhibits a high degree of stealth and obfuscation in cyberspace, making it difficult for Network Intrusion Detection System (NIDS) to ensure the accuracy and timeliness of detection. This paper researches machine learning and deep learning for intrusion detection in imbalanced network traffic. It proposes a novel Difficult Set Sampling Technique (DSSTE) algorithm to tackle the class imbalance problem. To verify the proposed method, we conduct experiments on the classic intrusion dataset NSL-KDD and the newer and comprehensive intrusion dataset CSE-CIC-IDS2018. We use classical classification models: random forest(RF), Support Vector Machine(SVM), XGBoost, MLP AlexNet, Mini-VGGNet.

Keywords: Machine Learning, Network traffic.

1.0 INTRODUCTION

With the development and improvement of Internet technology, the Internet is providing various convenient services for people. However, we are also facing various security threats. Network viruses, eavesdropping and malicious attacks are on the rise, causing network security to become the focus of attention of the society and government departments. Fortunately, these problems can be well solved via intrusion detection. Intrusion detection plays an important part in ensuring network information security. However, with the explosive growth of Internet business, traffic types in the network are increasing day by day, and network behavior characteristics are becoming increasingly complex, which brings great challenges to intrusion detection.How to identify various malicious network traffics, especially unexpected malicious network traffics, is a key problem that cannot be avoided. In fact, network traffic can be divided into two categories (normal traffics and malicious traffics). Furthermore, network traffic can also be divided into five categories: Normal, DoS (Denial of Service attacks), R2L (Root to Local attacks), U2R (User to Root attack) and Probe (Probing attacks). Hence, intrusion detection can be considered as a classification problem.

2. SYSTEM OVERVIEW

2.1 System architecture

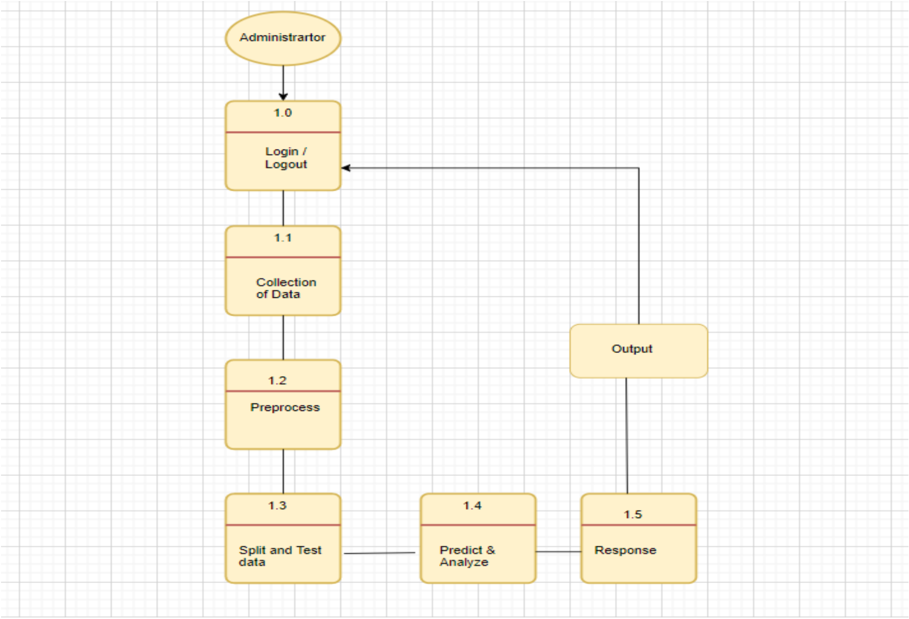
The three tier architecture is used when an effective distributed client/server design is needed that provides (when compared to the two tier) increased performance, flexibility, maintainability, reusability, and scalability, while hiding the complexity of distributed processing from the user. These characteristics have made three layer architectures a popular choice for Internet applications and net-centric information systems.

Existing System

The method of deep learning is to mine the potential features of high-dimensional data through training models and convert network traffic anomaly detection problems into classification problems. By training a large number of data samples, adaptive learning of the difference between normal behavior and abnormal behavior effectively enhances the real-time performance of intrusion processing.

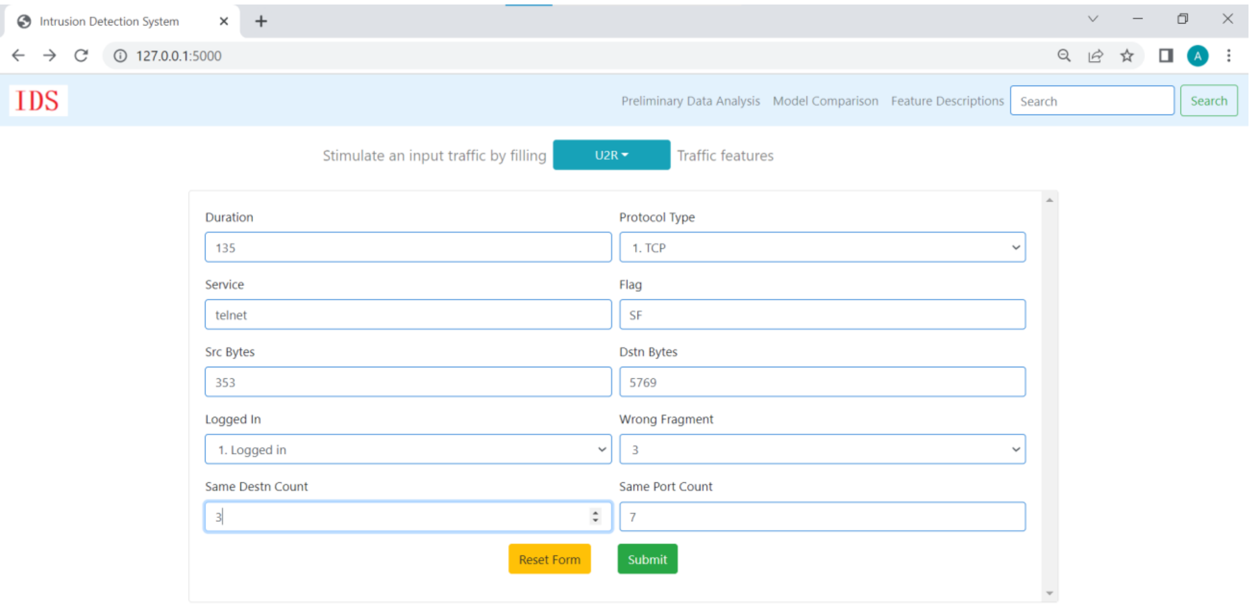
Proposed System

The usage of the classic NSL-KDD and the up-to-date CSECIC-IDS2018 as benchmark datasets and conduct detailed analysis and data cleaning. This work proposes a machine learning algorithm, reducing the majority samples and augmenting the minority samples in the difficult set, tackling the class imbalance problem in intrusion detection so that the classifier learns the differences better in training.**The** classification model uses Random Forest (RF), Support Vector Machine (SVM), XGBoost, we divide the experiment into 30 methods.



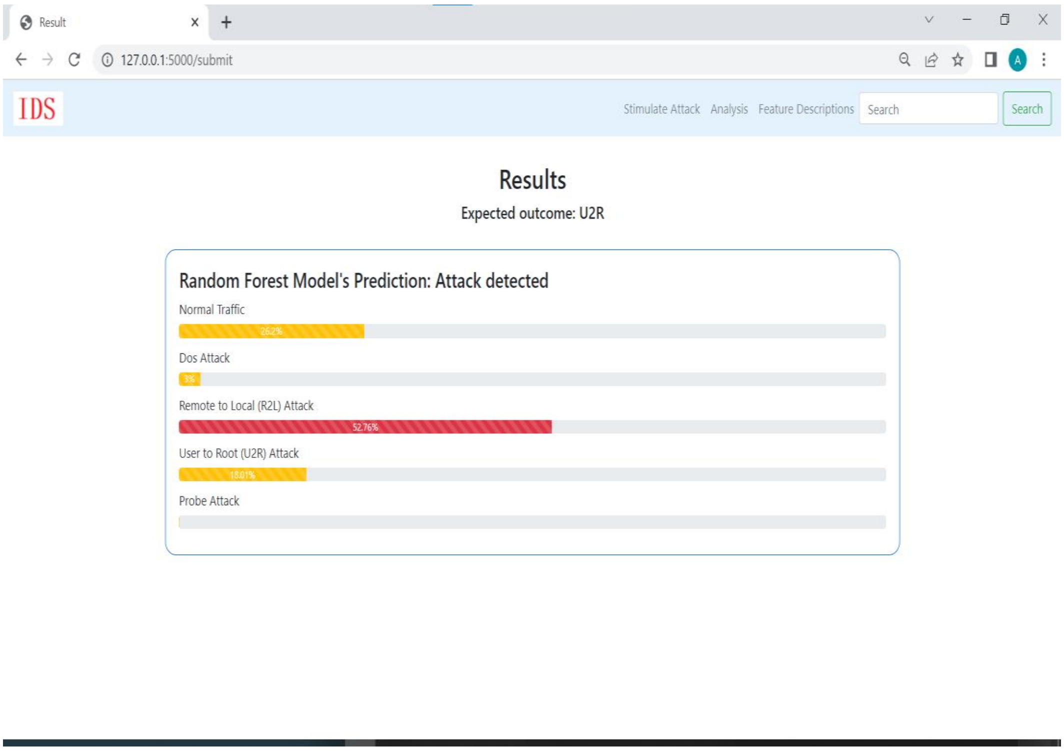
3. Results

3.1 Interface



Stimulate an input traffic by filling U2R Traffic features

Duration	Protocol Type
135	1. TCP
Service	Flag
telnet	SF
Src Bytes	Dstn Bytes
353	5769
Logged In	Wrong Fragment
1. Logged in	3
Same Destn Count	Same Port Count
3	7



Results

Expected outcome: U2R

Random Forest Model's Prediction: Attack detected

Normal Traffic 26.2%

Dos Attack 3%

Remote to Local (R2L) Attack 52.76%

User to Root (U2R) Attack 18.01%

Probe Attack

3.2 Detecting Network traffic with Accuracy

CONCLUSION

As network intrusion continues to evolve, the pressure on network intrusion detection is also increasing. In particular, the problems caused by imbalanced network traffic make it difficult for intrusion detection systems to predict the distribution of malicious attacks, making cyberspace security face a considerable threat. This paper proposed a novel Difficult Set Sampling Technique, which enables the classification model to strengthen imbalanced network data learning. A targeted increase in the number of minority samples that need to be learned can reduce the imbalance of network traffic and strengthen the minority's learning under challenging samples to improve the classification accuracy.

REFERENCES

[1]D. E. Denning, ‘‘An intrusion-detection model,’’ IEEE Trans. Softw. Eng., vol. SE-13, no. 2, pp. 222–232, Feb. 1987.

[2] N. B. Amor, S. Benferhat, and Z. Elouedi, ‘‘Naive Bayes vs decision trees in intrusion detection systems,’’ in Proc. ACM Symp. Appl. Comput. (SAC), 2004, pp. 420–424.

[3] M. Panda and M. R. Patra, ‘‘Network intrusion detection using Naive Bayes,’’ Int. J. Comput. Sci. Netw. Secur., vol. 7, no. 12, pp. 258–263, 2007.

[4] M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, ‘‘Support vector machine and random forest modeling for intrusion detection system (IDS),’’ J. Intell. Learn. Syst. Appl., vol. 6, no. 1, pp. 45–52, 2014.

[5] N. Japkowicz, ‘‘The class imbalance problem: Significance and strategies,’’ in Proc. Int. Conf. Artif. Intell., vol. 56, 2000, pp. 111–117.