MSc Project Proposal

Name: **Course:**
Supervisor (lecturer):

Current Modules (and previous modules if Computing or direct entrant)

The Project Title

Intrusion Detection System using Machine Learning

Project Introduction

Network Intrusions are the most significant threats that are used by cybercriminals to breach the data. This kind of network traffic is applied by cybercriminals to gain the access to the remote device through which they can steal sensitive and important information (Ahuja, et al., 2021). Cybercriminals penetrate the network and can access the organizational database and server to breach. So, when an organization or the organizational network will face such a kind of attack, a massive loss will be faced by them in terms of compromising sensitive data and confidential information (Ghouti & Imam, 2020). At the time the network will be under the intrusive attack, it is also merely prohibited as the users and the network administrator will lose access to the internet service and all other resources. So, the only way available to prohibit such an attack is by monitoring and capturing the network traffic and analyzing the network traffic data (Xia, et al., 2019). It will help to identify the suspected traffic in the network which will facilitate to design of the rule for the prohibition of network intrusions. In this research, machine learning algorithms will be applied for the detection of intrusive network traffics (Saberi, et al., 2019).

Problem Statement

Different kinds of intrusive attacks can be made on the network to breach data by gaining remote access to the user device and network. In most cases, cybercriminals use to make the detection system fool by imposing new features on the network traffic (Badugu, et al., 2019). The main reason behind this fact is the similarity of different types of Network traffic concerning the features. This makes the process of detecting intrusive traffic challenging. Additionally, features of certain intrusive traffic such as SYN Attacjk, SQL Injection etc. traffic behave like normal or benign which makes those untraceable unrecognized (Bhardwaj, et al., 2020). Using that loophole, cybercriminals may inject suspicious traffic into the network by utilizing those features. So, the main target of the research is to solve the challenge with the application of machine learning algorithms.

Literature Review

Saberi et al. (2019) have conducted their research to design the framework to detect Network intrusions in the cloud and IoT network. They have reviewed the previous research and have got the idea that most researchers have used the data from the CICD repository that provides sophisticated data in the field of cybersecurity. So, they have selected the dataset from the CICD repository namely the CICAndMal 2017 dataset that contains the traffic records of Network intrusions and benign traffic. They have faced the primary challenge to detect the important variables to decide the detection procedure of Network intrusions. To overcome the challenge, they have applied the correlation through

which they have selected the most important features out of the features enlisted in the dataset. After the selection of the data features, they have selected machine learning classifiers such as Random Forest Classifier, Naïve Bayes Classifier, Support Vector Machine and Artificial Neural Network. Finally, they have got the highest accuracy in using Artificial Neural Networks by 93% to detect Network intrusions.

Noorbehbahani & Saberi (2020) have proposed a model to detect Network intrusions traffic in the cloud network using machine learning. They have reviewed the previous research from which they have got the idea of the data selection and the application of machine learning to classify Network intrusions traffic. After getting a view of the data and the approaches, they have selected the dataset from the CICD website with the name CICAndMal 2017 dataset that contains the samples of malware that carried the Network intrusions agents in the cloud network. They have faced the primary challenge to detect the malware from the collected data samples. To overcome the challenge, they have selected the essential features which are important to identify Network intrusions. To classify Network intrusions traffic, they have selected machine learning classifiers such as Decision Tree Classifier and Support Vector Machine. Finally, they have classified the Network intrusions traffic using the Decision Tree Classifier with 92% accuracy.

Hirano & Kobayashi (2019) have researched the features of Network intrusions and have identified that there are different variants present in the cloud and IoT networks. To detect Network intrusions, they have focused on the existing models and have got an insight into the data and approaches. As Network intrusions are a kind of malware, they have selected the KDDCup dataset where a large variant of malware and Network intrusions are present. However, they have observed that there are more than 20 classes present from which they have to detect the Network intrusions. As the number of classes is higher, they have faced the challenge to detect Network intrusions from a large number of malware classes. To overcome the challenge, they have labelled the entire classes of the data into five labels and applied chosen classifiers to them. In this context, they have selected Random Forest, Support Vector Machine, and K-Nearest Neighbors. They have got the highest accuracy in the detection of Network intrusions using Random Forest by 92.51%.

Daku et al. (2018) have proposed a model based on machine learning to detect Network intrusions in the IoT and cloud networks. They have reviewed the previous models to get an idea about the approaches that have to be applied to detect Network intrusions and also have got an insight into the data. They have selected the CICAndMal 2017 dataset where 10 Network intrusions families are present. They have faced the challenge to handle a large number of types of network intrusions. To overcome the challenge, they have made the classes narrowed down to Network intrusions and benign and applied the selected classifiers such as Support Vector Machine, Neural Network and Random Forest Classifier. They have got the highest accuracy using Random Forest Classifier by 93.82% for the detection of network intrusions.

Method / Proposed Approach

Research Questions

The research questions for the project are chosen as follows:

1. Can the intrusive traffic be detected in the network with effectiveness?
2. Which network traffic features are more important to detect intrusive traffics?
3. Which machine learning classifier will be the best one to classify the intrusive traffic and will also show the significant improvement from the existing approaches?

Aim

The aim of the project is to detect network intrusion using a machine learning algorithm.

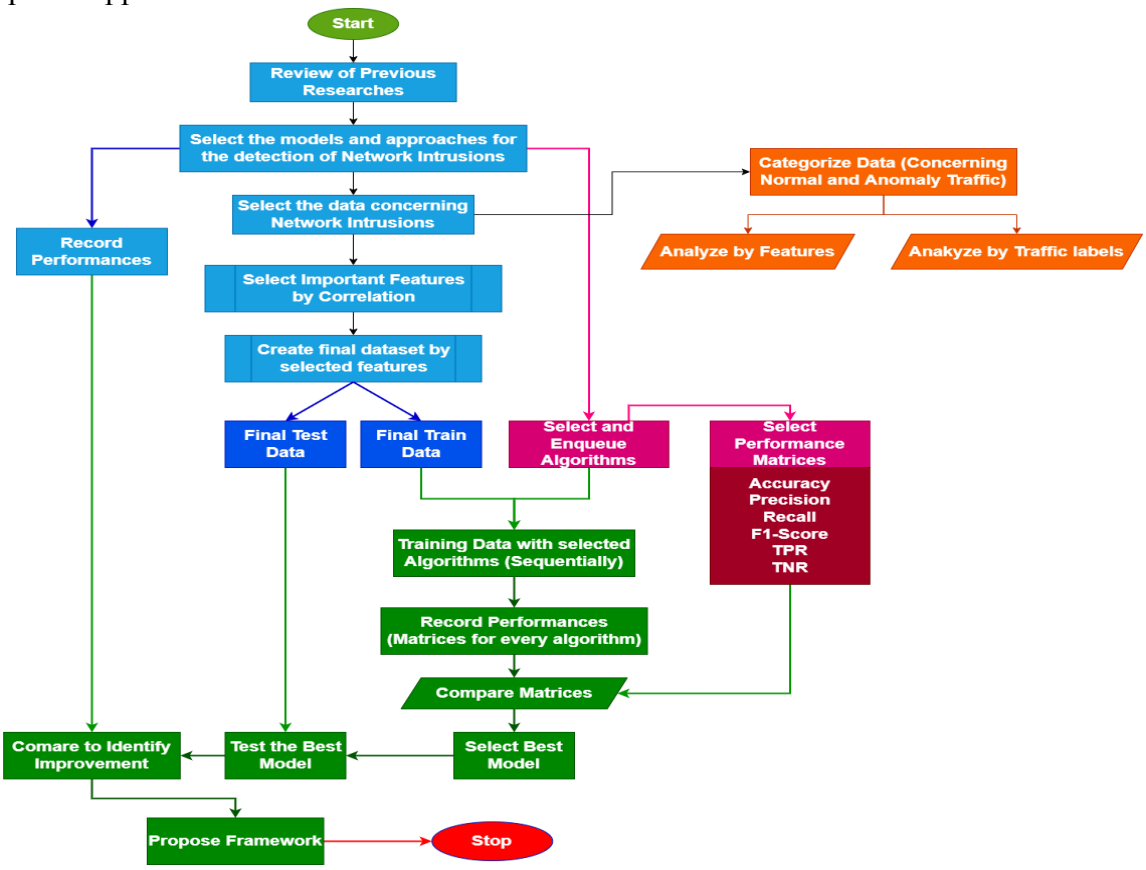
Objectives

The objectives of the project are as follows:

1. To understand the strength and weaknesses of a network and the way through which the intrusive traffic penetrates.
2. To review the previous approaches to identify the methods and techniques to detect network intrusions and other types of intrusive traffic in the network.
3. To select the machine learning models based on the previous research
4. To select the relevant data that will contain the records of the network traffic including normal traffic and intrusive traffic
5. To select the most suitable features from the data to use those as the final predictors
6. To apply the algorithms to the data and classify the network traffic along with finding the probability of the type of the network traffic (Monshizadeh, et al., 2019).
7. To compare the effectiveness of the models using classification metrices (Noorbehbahani & Saberi, 2020).
8. To compare the performance of the best model with the previous models and signify the improvement that can be seen in the present research compared to the previous models.

Proposed Approach

The proposed approach for the intrusion detection is shown below:



Tool Selection

Python 3 has been selected for the project for the detection of intrusive traffic. The main reasons behind the choice of Python are as follows:

a. Python is interpreted programming language
b. It is a dynamically typed programming language
c. It is open-source and thus no purchase or subscriptions are required to use it
d. It is portable and the same coding can be executed in any operating system.
e. Python programming can be done in any platform with the same piece of code and without changing the orientation and so it is referred to as the platform-independent programming language
f. Python has a large number of library support for machine learning and data analysis
g. Python programming language is free to use and it does not require any price to be paid for accessing and using it.

Technology selection

The technology that will be planned to be used in this project to fulfil the aim and objectives and to address the research questions are as follows:

i. Data Analytics
ii. Data Visualization
iii. Feature Engineering
iv. Machine Learning and Classification

Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Introduction, Aim and Objectives															
Review of existing models and study network intrusions and the features															
Select network intrusion data and algorithms															
Prepare the data by applying feature Selection and studying the features															
Train and Test the algorithms to network intrusions															
Compare the performances and select the most effective model															
Signify improvement by applying the model in the present research															
Concluding the research															

Potential Ethical or Legal Issues

The issues that are related to ethical and legal aspects are discussed below:

1. The research will not involve the device of any other person where the keylogging operation will be done.
2. The data on which the detection of network intrusions will be done should be collected from an open-licensed data vault
3. The programming tool will be selected such that it will be free to use and no need for payment or subscription for it.
4. No personal information will be collected during the project and study

The issues that are related to professional aspects are discussed below:

1. The research and execution of the project will be done professionally.

2. The project will be conducted so that the aim and the objectives will be fulfilled and the research questions can be addressed appropriately.
3. The referencing will be done as directed by university rules and regulations.

The issues that are related to social aspects are discussed below:

1. The research will help identify network intrusionss on the devices.
2. It will be helpful for the users to protect their computers and network from network intrusionss.

References

Abubakar, R. et al., 2020. An Effective Mechanism to Mitigate Real-Time DDoS Attack. *IEEE Access,* pp. 126215 - 126227.

Ahuja, N., Singal, G. & Mukhopadhyay, D., 2021. DLSDN: Deep Learning for DDOS attack detection in Software Defined Networking. *11th International Conference on Cloud Computing, Data Science & Engineering (Confluence),* pp. 1-5.

Alasmary, H. et al., 2019. Analyzing and Detecting Emerging Internet of Things Malware: A Graph-Based Approach. *Internet of Things Journal IEEE,* pp. 8977-8988.

Almolhis, N. & Haney, M., 2019. IoT Forensics Pitfalls for Privacy and a Model for Providing Safeguards. *Computational Science and Computational Intelligence (CSCI) 2019 International Conference,* pp. 172-178.

Badugu, M., Pusukuri, N. L. & Alluri, B. K. S. P. K. R., 2019. A Novel DDoS Detection Mechanism: Trust based Approach. *IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT),* pp. 1-6.

Bhardwaj, A., Mangat, V. & Vig, R., 2020. Hyperband Tuned Deep Neural Network With Well Posed Stacked Sparse AutoEncoder for Detection of DDoS Attacks in Cloud. *IEEE Access ,* pp. 181916 - 181929.

Bishnoi, S., Mohanty, S. & Sahoo, B., 2021. A Deep Learning-Based Methodology in Fog Environment for DDOS Attack Detection. *5th International Conference on Computing Methodologies and Communication (ICCMC),* pp. 135-139.

Daku, H., Zavarsky, P. & Malik, Y., 2018. Behavioral-Based Classification and Identification of Ransomware Variants Using Machine Learning. *17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE),* pp. 553-557.

Ghouti, L. & Imam, M., 2020. Malware classification using compact image features and multiclass support vector machines. *Information Security IET,* pp. 419-429.

Hirano, M. & Kobayashi, R., 2019. Machine Learning Based Ransomware Detection Using Storage Access Patterns Obtained From Live-forensic Hypervisor. *Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS),* pp. 1-5.

Iqbal, S. & Zulkernine, M., 2018. SpyDroid: A Framework for Employing Multiple Real-Time Malware Detectors on Android. *13th International Conference on Malicious and Unwanted Software (MALWARE),* pp. 167-171.

Jaiswal, M., Malik, Y. & Jaafar, F., 2018. Android gaming malware detection using system call analysis. *6th International Symposium on Digital Forensic and Security (ISDFS),* pp. 1-5.

Junyang Qiu, J. Z. W. L. L. P. S. N. Y. W. Y. X., 2019. A3CM: Automatic Capability Annotation for Android Malware. *Access IEEE,* Volume 7, pp. 147156-147168.

Monshizadeh, M. et al., 2019. Performance Evaluation of a Combined Anomaly Detection Platform. *IEEE Access,* pp. 100964 - 100978.

Noorbehbahani, F. & Saberi, M., 2020. Ransomware Detection with Semi-Supervised Learning. *10th International Conference on Computer and Knowledge Engineering (ICCKE),* pp. 1-5.

Saberi, M., Noorbehbahani, F. & Rasouli, F., 2019. Analysis of Machine Learning Techniques for Ransomware Detection. *16th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC),* pp. 182-186.

Singh, S., Fernandes, S. V., Padmanabha, V. & Rubini, P., 2021. MCIDS-Multi Classifier Intrusion Detection system for IoT Cyber Attack using Deep Learning algorithm. *Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV),* pp. 1-5.

Xia, S.-M.et al., 2019. A New Smart Router-Throttling Method to Mitigate DDoS Attacks. *IEEE Access,* pp. 107952 - 107963.

Yungaicela-Naula, N. M., Vargas-Rosales, C. & Perez-Diaz, J. A., 2021. SDN-Based Architecture for Transport and Application Layer DDoS Attack Detection by Using Machine and Deep Learning. *IEEE Access,* pp. 108495 - 108512.

Zhang, P. et al., 2021. Network-Wide Forwarding Anomaly Detection and Localization in Software Defined Networks. *IEEE/ACM Transactions on Networking,* pp. 1-7.