**ANOMALY BASED DETECTION IN SDN USING MACHINE LEARNING TECHNIQUES**

**APROJECTREPORT**

***Submittedby***

**Aravindan P-2019115018**

**Santhapravinraj E-2019115088**

**Bhairav AJ-2019115026**

*submittedtotheFacultyof*

## **INFORMATIONANDCOMMUNICATIONENGINEERING**

*inpartial fulfilment fortheawardofthedegree*

***of***

**BACHELOROFTECHNOLOGY**

***in***

**INFORMATIONTECHNOLOGY**



**DEPARTMENTOFINFORMATIONSCIENCEANDTECHNOLOGY COLLEGE OF ENGINEERING, GUINDY**

**ANNAUNIVERSITY CHENNAI 600 025**

**BONAFIDE CERTIFICATE**

Certified that this project report titled **Anomaly based detection in SDN using machine learning techniques** is the bonafide work of Aravindan P,Bhairav AJ and Santhapravinraj E who carried out project work under my supervision. Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

**PLACE:CHENNAI**

**DATE: 07.12.2022**  **Ms.G MAHALAKSHMI**

**PROJECT GUIDE**

**DEPARTMENT OF IST, CEG**

**ANNA UNIVERSITY**

**CHENNAI 600025**

**COUNTERSIGNED**

Dr. S SRIDHAR

HEAD OF THE DEPARTMENT

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY ANNA UNIVERSITY

CHENNAI 600025

**ABSTRACT**

The internet has caused a profound transformation in the globe. In actuality, it helps people keep up their social networks and connects them to other people in their networks when they need help. Sharing personal and professional information really entails a number of hazards for both individuals and corporations. Due to the importance of the internet in modern life, the security of our DATA is always under risk. IDS is crucial in defending internet users from malicious network assaults as a result. Intrusion Detection System (IDS) is a system that keeps an eye on network traffic for any unusual behavior and sends out alarms when it does.Malicious activity detection comes in two flavors: signature-based detection or misuse detection, where the IDS gathers data, analyses it, and then compares it to attack signatures kept in a large database. The second type of detection, known as an anomaly detection, considers any action that deviates from customary behavior to be malicious activity.

# **ACKNOWLEDGEMENT**

We wish to record our deep sense of gratitude and profound thanks to our project supervisor Ms.G.Mahalakshmi, Teaching Fellow, Department of Information Science and Technology, College of Engineering, Guindy for her keeninterest,inspiringguidance,constantencouragementwithourworkduring all stages, to bring this project into fruition.

WeareextremelyindebtedtoDr.SridharS,Professor,andHeadoftheDepartment ofInformationScienceandTechnology,AnnaUniversity,Chennai,forextending thefacilitiesoftheDepartmenttowardsourprojectandforhisunstintingsupport.

We would like to express our sincere thanks to the project committee members, Dr.J.INDUMATHI, Professor,Dr.S.SENTHIL KUMAR, Associate Professor,Dr.K.VIDYA, Assistant Professor, Dr. D.NARASHIMAN, Teaching Fellow,Department of Information Science and Technology, Anna University, Chennai for giving their valuable suggestions, encouragement and constant motivation throughout the duration of our project.

Aravindan P

Bhairav AJ

Santhapravinraj E

# **TABLEOFCONTENTS**

[ABSTRACT](#_bookmark0) iii

1. [INTRODUCTION](#_bookmark1) 1
   1. [BACKGROUND](#_bookmark2) 1
   2. [OBJECTIVE](#_bookmark3) 1
   3. [PROBLEMSTATEMENT](#_bookmark4) 1
   4. [ORGANIZATIONOFTHEREPORT](#_bookmark6) 2
2. [LITERATURESURVEY](#_bookmark7) 3
   1. [INTRODUCTION](#_bookmark8) 3
      1. [INTRUSION DETECTION USING SUPERVISED MACHINE LEARNING TECHNIQUES](#_bookmark9)

ANOMALY DETECTION USING MACHINE

LEARNING APPROACHES

SURVEY ON SDN BASEDNETWORK INTRUSION DETECTION SYSTEMUSING MACHINE LEARNING APPROACHES

1. [SYSTEMDESIGN](#_bookmark14) 5
   1. [INTRODUCTION](#_bookmark15) 5
   2. [TECHNICALARCHITECTURE](#_bookmark16) 5
   3. [DATAPREPROCESSING](#_bookmark18) 6

STANDARDISATION 6

* + 1. [NORMALISATION](#_bookmark19) 6

1. **IMPLEMENTATION**

RANDOM FOREST

**5REFERENCES**

**CHAPTER1**

### **INTRODUCTION**

### **1.1. BACKGROUND**

Intrusion is a set of actions that attempts to compromise the integrity, confidentiality, or availability of any resource on a computing platform. An intrusion detection system (IDS) is a system that monitors network traffic for doubtful activity and matters alert when such activity is exposed. While anomaly discovery and reportage is the primary function, some interruption detection systems are capable of taking actions when malicious activity or irregular traffic is detected, including blocking traffic sent from suspicious IP addresses. An IDS is a combination of hardware and software that detects intrusions in the network. It is used to track, identify and detect the intruders. It is a combination of both hardware and software that detects intrusions in the network. IDS is used to detect unauthorized intrusions that occur in computer systems and networks. Feature selection for intrusion detection is most important factor for the success of intrusion detection system.

### **1.2 OBJECTIVE**

The purpose of anomaly detection is to detect whether its attack or normal.

**1.3 PROBLEM STATEMENT**

An anomaly detection system should be able to identify all abnormal patterns and traffic using monitoring, detecting and responding to unauthorized activities within the system.

However, regarding its huge and unbalanced datasets,it encounters total data processing problem.Thus, different techniques have been presented which can handle this problem.

## **ORGANIZATION OF THE REPORT**

This thesis is organized into 6 chapters, describing each partof theprojectwithdetailedillustrationsandsystemdesigndiagrams.Thechaptersareasfollows:

**CHAPTER2**:Thischapterexplainstheliteraturesurveydetailsoftheexistingsystemswiththeirmethodologies,advantages,disadvantages,etc.

**CHAPTER 3**:Thischapterconsistsofthesystemdesignoftheprojectwiththeoverall architecture and the modules of the architecture and the descriptionofthemodules usedintheproject.

**CHAPTER 4:**

**CHAPTER 5:**

### **CHAPTER-2**

### **LITERATURE SURVEY**

**2.1 INTRUSION DETECTION USING SUPERVISED MACHINE LEARNING TECHNIQUES**

**2.1.1 Purpose**

This proposed modelfocus on the detecting the intrusion using the supervised machine learning algorithms

**2.1.2 Methodology**

The main objective of this paper is to present a survey for supervised learning algorithms and intrusion detection systems. We review necessary concepts that are related to IDS. The concept of IDS is first presented, where we discuss its definition, types, and importance. Then,review popular supervised learning algorithms and popular data sets in this domain and the concept of dimensionality reduction.

**2.2 ANOMALY DETECTION USING MACHINE LEARNING APPROACHES**

**2.2.1 Purpose**

The results of this proposed work was to enhance the performance of detecting anomalies with a low time frame.

**2.2.2 Methodology**

A hybrid Naïve Bayes-SVM model (NB-SVM) has been considered here that has two layers for classification. Naïve Bayes works well in real-time and consumes less time in computation; it is chosen as the first classifier. Initially, the incoming data or the anomalous data were first fed through the NB classifier, and the output is generated. The intermediate result is then again fed as input data to SVM. After going through the computation process of SVM, the final output is achieved. The final output differentiates between a normal attack and an anomaly

**2.3 SURVEY ON SDN BASEDNETWORK INTRUSION DETECTION SYSTEMUSING MACHINE LEARNING APPROACHES**

**2.3.1 Purpose**

In this proposed model, they reviewed various recent works on machine learning (ML) methods that leverage SDN to implement detection system.

**2.3.2 Methodology**

In this proposed model, they have provided an overview of programmable networks and examined the emerging field of Software-Defined Networking (SDN). We also outlined various intrusion detections mechanisms using ML/DL approaches. We emphasized software-defined networking (SDN) technology as a platform using ML/DL approaches to detect vulnerabilities and monitor networks.

### **CHAPTER 3**

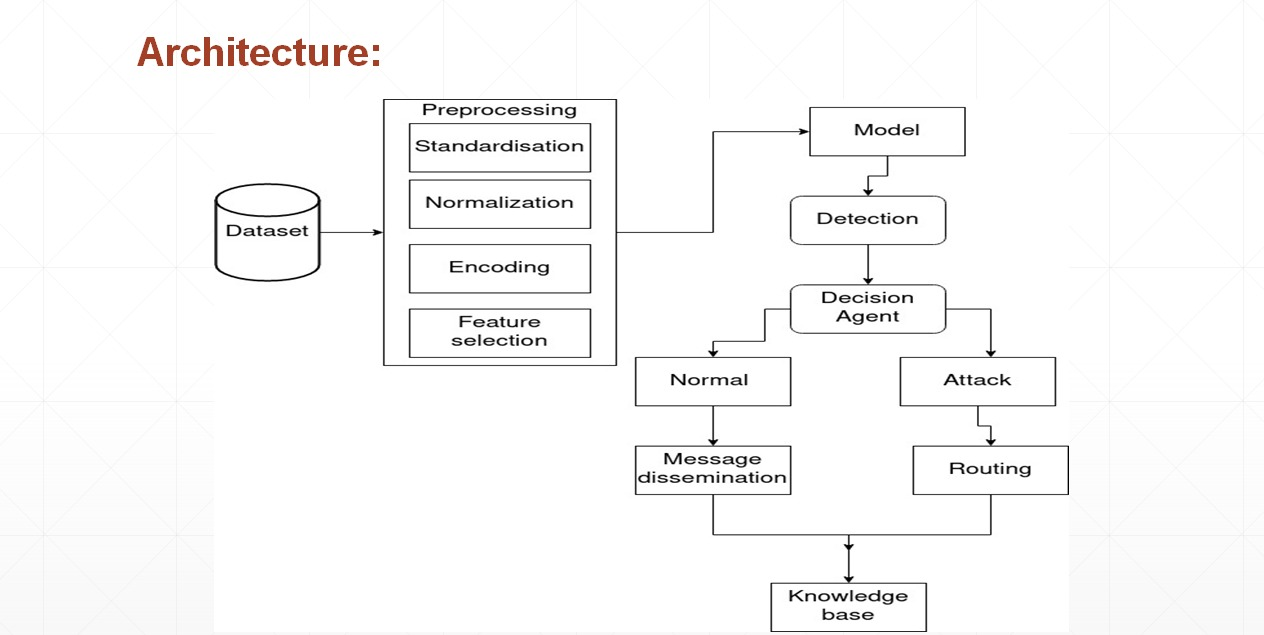
### **SYSTEM DESIGN**

### **3.1 INTRODUCTION**

### This chapter consists of the system design of the project with the technical architecture

### **3.2 TECHNICAL ARCHITECTURE**

pls do this

****

**3.3 Data pre-processing**

**3.3.1 Standardization**

Standardization entails scaling data to fit a standard normal distribution. A standard normal distribution is defined as a distribution with a mean of 0 and a standard deviation of 1. Visualizing standardization. To better understand standardization, it would help to visualize its effects on some data.

**3.3.2 Normalization**

Normalization is a scaling technique in Machine Learning applied during data preparation to change the values of numeric columns in the dataset to use a common scale. It is not necessary for all datasets in a model. It is required only when features of machine learning models have different ranges.

**3.3.3 One hot encoding**

One hot encoding can be defined as the essential process of converting the categorical data variables to be provided to machine and deep learning algorithms which in turn improve predictions as well as classification accuracy of a model.

**CHAPTER 4**

**IMPLEMENTAION**

**PRINCIPAL COMPONENT ANALYSIS:**

 The steps to perform PCA are the following:

Step 1: Standardize the data.

Step 2:Compute the covariance matrix of the features from the dataset.

Step 3:Perform eigendecompositon on the covariance matrix.

Step 4:Order the eigenvectors in decreasing order based on the magnitude of their corresponding eigenvalues.

Step 5:Determine k, the number of top principal components to select.

Step 6:Construct the projection matrix from the chosen number of top principal components.

Step 7: Compute the new k-dimensional feature space.

**Random forest:**

**Step 1** − First, start with the selection of random samples from a given dataset.

**Step 2** − Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.

**Step 3** − In this step, voting will be performed for every predicted result.

**Step 4** − At last, select the most voted prediction result as the final prediction result.



**CHAPTER 5**

**REFERENCES**

[1] Emad E. Abdallah, Wafa’ Eleisah, Ahmed Fawzi Otoom,”Intrusion detection using supervised machine learning techniquesThe 13th International Conference on Ambient Systems, Networks and Technologies” (ANT) March 22-25, 2022, Porto, Portugal

[2]Ahmed, Md. Rayhan, Islam, salekul, Shatabda, Swakkhar; Islam, A. K. M. Muzahidul, Robin, Md. TowhidulIslam (2021): Intrusion Detection System in Software-Defined Networks Using Machine Learning and DeepLearning Techniques.

[3]Mausumi Das, Nath Tappalina ,” Anomaly Detection Using Machine Learning Approaches” Azerbaijan Journal of High Performance Computing, Vol 3, Issue 2, 2020

[4] Nasrin Sultana,NaveenChilamkurti,Wei Peng , RabeiAlhadad,” Survey on SDN based network intrusion detection system using machine learning approaches”12 January 2018 Springer ScienceBusiness Media, LLC, part of Springer Nature 2018

[5] Nabila Farnaaz M. A. Jabba “Random Forest Modeling for Network Intrusion Detection System”, Twelfth International Multi-Conference on Information Processing-2016