KWARA STATE UNIVERSITY MALETE

NAME
Gbadamosi Sodiq Alabi

MATRIC NUMBER
20D/47CS/01255

COURSE CODE
CSC 499

PROJECT TITTLE
MACHINE LEARNING IN INTRUSION DETECTION SYSTEM

LECTURER IN CHARGE
Dr. S.O AbdulSalam

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Background of the Study:

With the increase in the number of cyberattacks, the need for effective security systems to protect computer networks has become essential. One of the critical components of security systems is intrusion detection systems (IDS), which are designed to detect potential attacks on computer networks.

An IDS is a software application that monitors network traffic for suspicious activity and raises an alert if it detects any potential intrusion. Intrusions can come in various forms, such as malware, viruses, phishing attacks, and Denial of Service (DoS) attacks. IDS can detect intrusions by analyzing network traffic patterns, anomalies, and signatures of known attacks.

Traditionally, IDS were rule-based systems that used a set of pre-defined rules to identify potential attacks. However, these systems were limited in their ability to detect unknown attacks and were vulnerable to evasion techniques used by attackers.

Machine learning has emerged as a powerful technique for developing IDS, which can improve the detection accuracy and reduce the false-positive rate of traditional rule-based systems. Machine learning algorithms can learn from the data and identify patterns that are indicative of potential attacks. The advantage of machine learning-based IDS is their ability to learn and adapt to new attack types without the need for manual rule definition.

Several machine learning algorithms have been used for developing IDS, including decision trees, neural networks, support vector machines, and ensemble methods. One of the popular algorithms used for IDS is the Random Forest algorithm, which is an ensemble learning method that combines multiple decision trees to improve the accuracy and performance of the system.

In this study, we propose a network-based intrusion detection system that uses the Random Forest algorithm for detecting potential attacks. The system will be developed and evaluated using a large dataset of network traffic data, with the goal of achieving a high detection rate with a low false-positive rate and fast response times.

The proposed system will be developed using the Python programming language, which is a popular language for machine learning applications. The Scikit-learn library will be used to implement the Random Forest algorithm. The system will be trained on a large dataset of network traffic data, which will be preprocessed to remove noise and irrelevant information.

The system will be evaluated based on its detection accuracy, false-positive rate, and response time. The evaluation will be conducted using a simulated network environment to test the effectiveness of the system in detecting potential attacks. The study's findings are expected to contribute to the development of advanced security systems that can protect organizations from emerging cyber threats.

Statement of the Problem:

The existing IDSs face several shortcomings, such as high false-positive rates, slow response times, and difficulty in detecting unknown threats. These limitations can result in missed attacks and an increased workload for security teams. To address these shortcomings, a more advanced IDS that leverages machine learning algorithms can be developed. Such an IDS can reduce false positives, increase detection accuracy, and improve response times to potential attacks.

The proposed solution involves using machine learning algorithms to analyze network traffic and identify patterns that indicate malicious activity. The system will be trained on large datasets of known attacks to ensure it can accurately detect potential threats. The proposed IDS will improve the efficiency and accuracy of the security system, enabling quicker responses to potential threats and reducing the workload on security teams.

Aim and Objectives:

The aim of this project is to develop an intrusion detection system that uses machine learning algorithms to detect potential security threats accurately. The objectives of the project are:

i. To collect and preprocess a large dataset of both benign and malicious traffic data.

ii. To develop and train a machine learning model on the dataset.

iii. To deploy the model to monitor the network and detect potential attacks in real-time.

iv. To evaluate the performance of the system and compare it to existing IDSs.

Significance of the Study:

The proposed IDS has several benefits, including reducing the workload on security teams, increasing detection accuracy, and improving response times to potential attacks. The IDS can also help prevent data breaches and minimize the impact of successful attacks. Furthermore, the project can contribute to the development of advanced security systems that can protect organizations from emerging cyber threats.

Scope of the study:

The proposed intrusion detection system will be developed and evaluated using the following:

- Area of application: The system will be developed for network-based intrusion detection, which involves analyzing network traffic to detect potential attacks.
- Dataset used: A large dataset of network traffic data will be collected from various sources, both benign and malicious, for training and testing the machine learning model.
- Data processor used: The collected data will be preprocessed to remove noise and irrelevant information. Feature extraction techniques will be applied to identify the relevant features of the dataset.
- ML algorithm used: The Random Forest algorithm will be used for the development of the machine learning model, which is known for its high accuracy and performance.
- Intended tools used in implementing: Python will be used to develop the intrusion detection system, with the Scikit-learn library used for the implementation of the Random Forest algorithm.
- Performance: The performance of the system will be evaluated based on its detection accuracy, false-positive rate, and response time. The system will be tested in a simulated network environment to evaluate its effectiveness in detecting potential attacks. The goal is to achieve a high detection rate with a low false-positive rate and fast response times.