A Synopsis of

Minor Project [CC3270]

Cybersecurity Threat Prediction System

Submitted to Manipal University Jaipur

towards the partial fulfillment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer and Communication Engineering

Jan – May 2025

By:

Tuhina Rath

23FE10CCE00064

Under the guidance of:

Dr. Vijay Shankar Sharma

Department of Computer and Communication Engineering

Manipal University Jaipur

Introduction

The rapid digital transformation across industries has significantly increased the threat landscape of cyberattacks, making cybersecurity a critical concern. The growing reliance on digital infrastructure has made organizations more vulnerable to cyber threats such as data breaches, malware, phishing, and denial-of-service attacks. The consequences of these attacks range from financial losses to reputational damage, making cybersecurity a top priority for businesses and governments worldwide.

Traditional Intrusion Detection Systems (IDS) rely on signature-based techniques, which require constant updates and struggle to detect new, evolving threats. While these systems have been effective in identifying known attack patterns, they fail to adapt to sophisticated and previously unseen threats, such as zero-day exploits and AI-driven cyberattacks. The limitations of traditional IDS have necessitated the development of more advanced and adaptable threat detection solutions.

This project aims to build an AI-powered Cybersecurity Threat Prediction System, which integrates machine learning (ML) and cloud computing to detect, classify, and predict cyber threats in real-time. By analyzing network traffic, detecting suspicious patterns, and providing proactive alerts, the system enhances security response efficiency and reduces the impact of cyberattacks. The combination of AI and cloud technologies ensures scalability, faster processing, and improved accuracy in identifying emerging cyber threats.

Motivation

The increasing volume and sophistication of cyberattacks demand advanced, data-driven approaches to threat detection. Organizations and individuals face continuous security risks, with cybercriminals leveraging increasingly complex attack methods. Traditional **rule-based security solutions** are no longer sufficient to handle modern cybersecurity challenges, as they depend on predefined attack signatures and require frequent updates to remain effective.

One of the major challenges with **existing IDS solutions** is their high false positive rates, which lead to unnecessary alerts and inefficiencies in security operations. Additionally, signature-based systems struggle to detect novel threats that do not match existing patterns. These shortcomings highlight the need for AI-driven cybersecurity solutions capable of learning from historical attack data and identifying emerging threats with greater accuracy.

By integrating ML-driven anomaly detection with cloud-based deployment, this project aims to enhance cybersecurity monitoring with higher accuracy, real-time threat detection, and predictive capabilities. The ability to proactively detect and respond to cyber threats will not only strengthen security defences but also provide a scalable and adaptable solution suitable for modern digital environments.

Problem Statement

Cybersecurity threat detection has traditionally been performed using signature-based IDS**,** which involves matching predefined attack signatures to detect threats. However, this approach has several limitations. Signature-based IDS struggles to detect zero-day attacks and evolving cyber threats, leading to a reactive rather than proactive security strategy. Furthermore, the high reliance on known attack patterns means new and sophisticated threats often go unnoticed, leaving organizations vulnerable to advanced persistent threats (APT) and AI-driven cyberattacks.

Most existing IDS solutions also face challenges in handling large-scale, real-time network traffic. High false positive rates often result in security teams being overwhelmed with alerts, many of which are benign, leading to inefficient threat response. Additionally, signature-based systems require constant manual updates, making them difficult to maintain in dynamic network environments.

To address these challenges, this project aims to develop an AI-powered cybersecurity threat prediction system that integrates machine learning and cloud computing for more effective threat detection. The objectives of this project are:

1. **Enhancing cybersecurity threat detection accuracy** by incorporating **supervised and unsupervised ML models** to identify both known and unknown threats.
2. **Reducing false positive rates** using anomaly detection techniques.
3. **Deploying a scalable, cloud-based solution** that can process and analyze network traffic in real time.
4. **Providing an intelligent threat monitoring dashboard**, enabling security teams to visualize and respond to threats efficiently.

Pros & Cons of Existing Models

Pros:

- Signature-based IDS is well-established and widely used in cybersecurity.
- Anomaly-based detection captures unknown attack patterns, improving detection capabilities.
- AI-powered security models continuously learn from new threats, enhancing adaptability.

Cons:

- Signature-based models fail to detect novel or zero-day threats.
- Traditional IDS has **high false positive rates**, making it difficult to distinguish real threats.
- Real-time cybersecurity monitoring requires significant computational resources and cloud integration.

By leveraging machine learning algorithms and cloud technologies, this project seeks to overcome these limitations, providing a more effective and scalable cybersecurity solution than traditional IDS.

Methodology/Planning of Work

Methodology

1. **Data Collection**: Use cybersecurity datasets like CICIDS 2017, KDD Cup 1999, and UNSW-NB15 to gather normal and malicious network traffic data. Generate simulated attacks for testing if needed.
2. **Preprocessing**: Clean and organize the data by removing errors, normalizing values, and selecting important features like packet size and connection duration.
3. **Feature Extraction**: Use machine learning techniques to detect patterns in network traffic. Supervised models will classify known threats, while unsupervised models will detect unknown attacks.
4. **Model Training**: Train different ML models (e.g., Random Forest, SVM) to recognize cyber threats. We may also test deep learning models for better accuracy.
5. **Evaluation & Testing**: Check how well the models perform using accuracy, precision, recall, and F1-score. Improve models using cross-validation.
6. **Integration & Deployment**: Create a real-time dashboard to monitor threats. Deploy the system on AWS or Google Cloud for scalability and efficiency.

Planning of Work (Gantt Chart Overview)

Task	**Week 1-2**	**Week 3-4**	**Week 5-6**	**Week 7-8**	**Week** 9-10	**Week 11-12**	**Week 13-14**	**Week 15**
Literature Survey								
Data Collection								
Preprocessing								
Feature Extraction								
Model Training								
Testing & Evaluation								
Deployment & Documentation								

Weeks 1-2 (Literature Survey & Requirement Analysis)

- Define project goals and key objectives.
- Research existing cybersecurity threat detection systems and machine learning approaches.
- Identify limitations in traditional Intrusion Detection Systems (IDS) and how AI/ML can improve detection.

Weeks 3-4 (Data Collection & System Design)

- Gather cybersecurity datasets (**CICIDS 2017, KDD Cup 1999**) and preprocess raw network traffic data.
- Design system architecture for integrating **ML models** with cloud deployment.
- Plan a database schema to store network logs and threat classifications.
- Outline UI/UX wireframes for a threat monitoring dashboard.

Weeks 5-6 (Preprocessing & Feature Extraction)

- Clean and normalize network traffic data to remove inconsistencies.
- Extract key features like **packet size, connection duration, and request frequency**.
- Apply feature selection techniques to improve model efficiency.

Weeks 7-8 (Model Training & Integration)

- Train individual **supervised ML models** (Random Forest, SVM) to classify threats.
- Implement **unsupervised models** (Isolation Forest, One-Class SVM) for anomaly detection.
- Experiment with **deep learning techniques** (e.g., LSTMs) to improve accuracy.

Weeks 9-10 (Testing & Evaluation)

- Evaluate model performance using **accuracy, precision, recall, and F1-score**.
- Compare different ML models and select the best-performing approach.
- Optimize model parameters to reduce false positives and improve real-time detection.
-

Weeks 11-12 (Deployment & Documentation)

- Deploy the trained model on a **cloud platform** (AWS/Google Cloud) for scalability.
- Develop a real-time **web dashboard** to visualize detected threats.
- Document system architecture, model performance, and key findings for future improvements.

Weeks 13-14 (Final Testing & Refinement)

- Conduct stress testing on real-time network traffic data.
- Integrate user feedback and fine-tune model performance.
- Optimize system latency for faster threat detection.

Week 15 (Final Deployment & Maintenance)

- Finalize system deployment and ensure stable operations.
- Set up a monitoring and maintenance protocol for continuous improvements.

Facilities Required for Proposed Work

Software Requirements:

- **Operating System** – Windows/Linux/macOS
- **Programming Language** – Python
- **ML Libraries** – Scikit-learn, TensorFlow/PyTorch
- **Cloud Platforms** – AWS, Google Cloud (for deployment)
- **Web Framework** – Flask for dashboard

Hardware Requirements:

- **Processor** – Intel i5/i7 or AMD Ryzen 5/7
- **RAM** – Minimum 8GB (Recommended: 16GB)
- **Storage** – SSD with at least 256GB

Bibliography/References

1. Abbasi, A., et al. (2020). "AI-Driven Intrusion Detection Systems for Cybersecurity." *IEEE Transactions on Cybersecurity*.
2. Tariq, S., et al. (2019). "Anomaly-Based Cyber Threat Detection Using Machine Learning." *Journal of Cybersecurity Research*.
3. Kumar, R., et al. (2022). "A Comparative Study of Machine Learning Techniques for Network Intrusion Detection." *Springer Security Analytics*.
4. MITRE ATT&CK Framework (https://attack.mitre.org/) – A threat intelligence repository.
5. CICIDS 2017 Dataset (https://www.unb.ca/cic/datasets/ids-2017.html) – Used for training cybersecurity ML models.