Network Traffic Analysis Using Machine Learning

BITS ZC4999T: Capstone Project

by

Rushil Gupta

202117bh082

Capstone Project work carried out at

HCL Tech, Lucknow

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI (RAJASTHAN)

July,2025

Birla Institute of Technology & Science, Pilani
Work-Integrated Learning Programmes Division

BITS ZC499T Capstone Project Outline

ID No.	: 202117bh082
NAME OF THE STUDENT	: Rushil Gupta
EMAIL ADDRESS	: 202117bh082@wilp.bits-pilani.ac.in
STUDENT'S EMPLOYING ORGANIZATION & LOCATION	: HCL Tech, Lucknow
MENTOR'S NAME	: Mohammed Mutheeb Parveez
MENTOR'S EMPLOYING ORGANIZATION & LOCATION	: HCL Tech, Bangalore
MENTOR'S EMAIL ADDRESS	: mohd.parveez@hcltech.com
CAPSTONE PROJECT TITLE	: Network Traffic Analysis Using Machine Learning

1. Summary:

This project, titled **Network Traffic Analysis using Machine Learning**, aims to develop an intelligent Intrusion Detection System (IDS). The system will leverage machine learning algorithms to analyze network packet logs near-real-time to identify anomalous activities and potential cyberattacks. By first establishing a baseline of normal network behavior using unsupervised learning, the model will be capable of detecting novel or "zero-day" threats that signature-based systems might miss. Further, a supervised classification model will be trained to categorize detected anomalies into specific attack types such as DDoS, Brute Force, and Port Scanning, and assess their severity. The outcome will be an automated system that can detect intrusions, classify them, and generate alerts for suspicious traffic patterns, enhancing network security and situational awareness. The core technologies will include Python, with libraries such as Scikit-learn, TensorFlow, and Pandas for data manipulation and modeling.

2. Background:

In the current digital landscape, the frequency, scale, and sophistication of cyberattacks are constantly increasing. Traditional security measures, like firewalls and signature-based Intrusion Detection Systems, are often reactive and struggle to keep pace with evolving attack vectors. They primarily rely on predefined rules and signatures of known threats, leaving them vulnerable to new and unknown attacks. This project addresses this gap by applying machine learning to network security. By learning patterns directly from traffic data, an ML-based IDS can create a dynamic baseline of "normal" behavior. Any significant deviation from this baseline is flagged as a potential anomaly. This approach, particularly using unsupervised learning, enables the detection of previously unseen threats. This project will demonstrate the practical application of data science in cybersecurity, a critical skill set for addressing modern security challenges.

3. Objective:

- To develop a robust data processing pipeline for ingesting, cleaning, and transforming raw network traffic data into a format suitable for machine learning analysis.
- To implement and train unsupervised machine learning models (e.g., Isolation Forest, Autoencoders) to effectively distinguish between normal and anomalous network traffic.
- To build a multi-class classification model capable of accurately identifying the type of detected attack (e.g., DDoS, Port Scan, Brute Force).
- To design a system for assessing the severity of detected threats based on their characteristics.
- To create a proof-of-concept alert system that notifies administrators of suspicious activities and their potential impact.

4. Scope of Work:

- Data Acquisition and Preparation: Obtain a suitable public network traffic dataset. Perform Exploratory Data Analysis (EDA), data cleaning, normalization, and feature engineering to extract meaningful features from packet and flow data.

- Unsupervised Anomaly Detection: Implement, train, and evaluate various unsupervised learning algorithms to establish a baseline of normal traffic and detect deviations.

- Supervised Attack Classification: Develop a supervised learning model using a labeled portion of the dataset. This model will be trained to classify detected anomalies into specific, predefined attack categories.

- Model Evaluation: Rigorously evaluate the performance of both unsupervised and supervised models using appropriate metrics such as Precision, Recall, F1-Score, Accuracy, and Confusion Matrices.

- Severity and Alerting System: Design and implement a module that assigns a severity score to detected threats and triggers alerts for high-severity events.

- Documentation and Reporting: Maintain detailed documentation of the project methodology, code, experiments, and results, culminating in a comprehensive final report.

5. Plan of work:

Week	Activity
1-2	Research existing ML-based IDS approaches. Select and acquire a suitable network dataset. Set up the Python development environment with necessary libraries.
3-4	Perform detailed Exploratory Data Analysis (EDA) to understand the dataset's structure and features. Begin data cleaning, handling missing values, and preparing the initial data processing pipeline.

5-7	Engineer relevant features for detecting attacks. Implement and train initial unsupervised models for general anomaly detection. Evaluate their ability to separate normal traffic from attacks.
8-10	Label the data for supervised learning. Implement and train multi-class classification models to identify specific attack types.
11-12	Perform rigorous evaluation of all models using test set. Fine-tune the hyperparameters of the best-performing models to optimize performance. Analyze results using confusion matrices and ROC curves.
13	Develop the logic for assigning a severity level to detected threats. Implement a prototype alerting mechanism.
14	Integrate the data pipeline, anomaly detection model, classification model, and alerting system. Test the entire workflow on unseen data to ensure functionality.
15	Create a basic IDS tool containing the model.
16	Prepare documentation for final report and mentor's review.

6. Literature References:

Dataset:

- Kaggle - Malware Detection in Network Traffic Data

Technical Documentation:

- Scikit-learn Documentation: https://scikit-learn.org/stable/documentation.html
- Pandas Documentation: https://pandas.pydata.org/docs/
- TensorFlow Tutorials: https://www.tensorflow.org/tutorials

Key Research Papers & Books:

- Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization." *ICISSp*, 1, 108-116. (Paper describing the CIC-IDS2017 dataset).
- Chandola, V., Banerjee, A., & Kumar, V. (2009). "Anomaly detection: A survey." *ACM computing surveys (CSUR)*, 41(3), 1-58.
- Kurose, J. F., & Ross, K. W. (2021). *Computer Networking: A Top-Down Approach*. Pearson. (For fundamental networking concepts).

7. Remark from mentor:

Rushil's proposal for an ML-powered Network Traffic Analyzer effectively bridges the gap between theoretical machine learning concepts and their practical application in solving real-world cybersecurity problems. The scope, which includes building an end-to-end pipeline from data ingestion to an automated alert system, is comprehensive. The focus on unsupervised learning for detecting novel attacks is particularly noteworthy and reflects a mature understanding of the domain's challenges. The project plan is structured logically, and I believe Rushil possesses the necessary skills to execute it successfully.
I approve this project for dissertation submission.

8. Mentor & Examiner Details:

Mentor:

Name: Mohammed Mutheeb Parveez
Designation: Group Manager
Organization: HCL Tech
Email: mohd.parveez@hcltech.com
Phone: +919902395353

Additional Examiner:
Name: Kalash Dutt
Designation: Senior Software Engineer
Organization: HCL Tech
Email: kalash.dutt@hcltech.com
Phone: +918565968856

Consent for Digital Signature – BITS WILP

Signature of Student	**Signature of Mentor**	**Signature of Additional Examiner**
Name: Rushil Gupta	Name: Mutheeb Parveez	Name: Kalash Dutt

Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division

BITS ZC499T: Capstone Project EC-1: Capstone Project Outline Evaluation Sheet

ID No. : 202117bh082

NAME OF THE STUDENT : Rushil Gupta

EMAIL ADDRESS : 202117bh082@wilp.bits-pilani.ac.in

STUDENT'S EMPLOYING ORGANIZATION & LOCATION : HCL Tech, Lucknow

MENTOR'S NAME : Mohammed Mutheeb Parveez

MENTOR'S EMPLOYING ORGANIZATION & LOCATION : HCL Tech, Bangalore

PROPOSED PROJECT TITLE : Network Traffic Analysis Using Machine Learning

---

CAPSTONE PROJECT OUTLINE EVALUATION
(Please put a tick (✔) mark in the appropriate box)

EC No.	Component	Excellent	Good	Fair	Poor
1.	Capstone Project Outline	✓			

	Mentor	Additional Examiner
Name	Mohammed Mutheeb Parveez	Kalash Dutt
Qualification	MBA	BCA
Designation	Group Manager	Senior Software Engineer
Employing Orgn and Location	HCL Tech, Bangalore	HCL Tech, Noida
Phone No. (with STD Code)	+919902395353	+918565968856
Email Address	mohd.parveez@hcltech.com	kalash.dutt@hcltech.com
Signature	Mohd Parveez	Kalash Dutt
Date	04-Aug-2025	04-Aug-2025