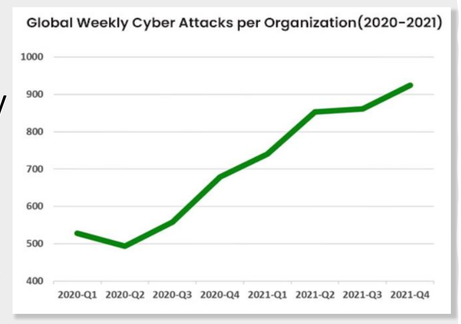
Anomaly Detection in Network Traffic for Security

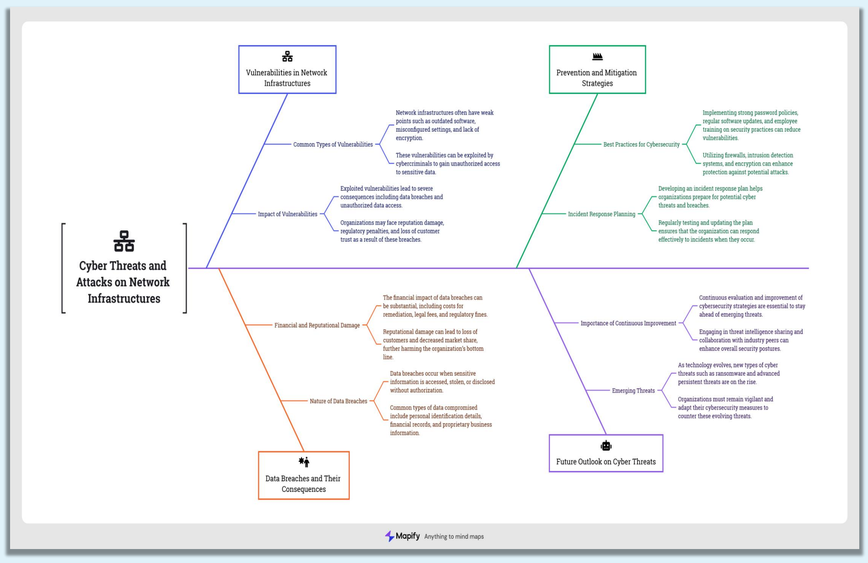
Under the supervision of
Dr. Maha Amin Hassanein
MTH 2253

Faculty of Engineering
Cairo University

Problem Statement

With the increasing reliance on the internet for communication, data storage, and online services, network security has become a critical concern. Cyber threats and attacks exploit vulnerabilities in network infrastructures , leading to data breaches, loss of sensitive information, and financial losses.





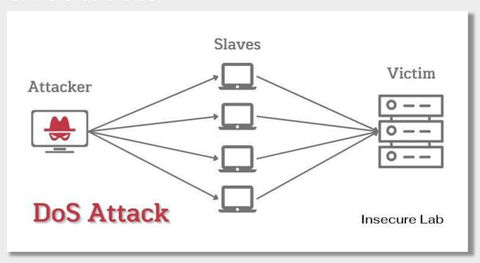
Research Objectives

This project aims to:

1.Accurately identify abnormal traffic patterns using statistical and machine learning techniques.
2.Develop a model that differentiates between normal and malicious network traffic.
3.Implement a predictive approach to detect anomalies in real-time, serving as an early warning system.
4.Minimize false positives while maintaining high detection accuracy.

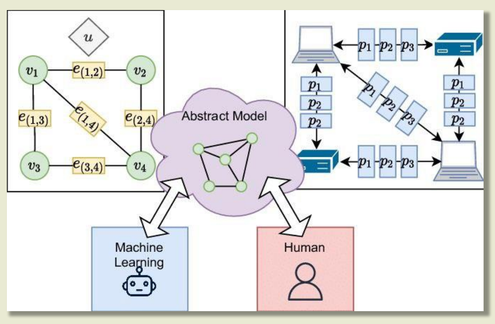
Background

Anomaly detection in network traffic is essential for identifying unusual patterns that may indicate security risks, such as DoS attacks, infiltration, or unauthorized access. With large, complex datasets , distinguishing between normal traffic and malicious attacks is a significant challenge.



Motivation

As internet-based systems become more integral to daily operations, preventing and mitigating cyber threats is essential for maintaining data confidentiality, integrity, and availability. Effective anomaly detection models can provide early warnings of security threats, reducing the risk of data breaches and unauthorized access.



Methodology

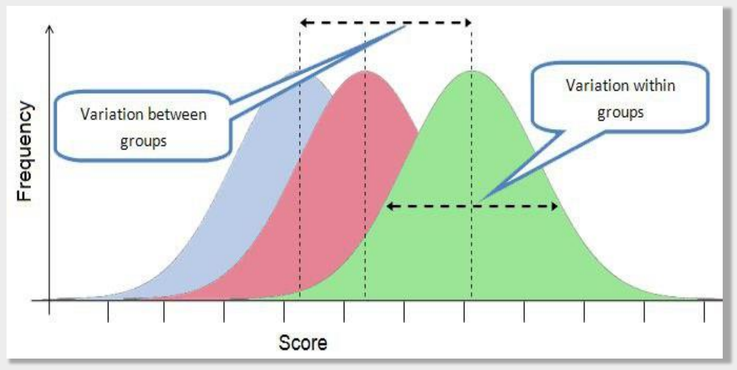
Research Design: This project use a machine learning approach to analyze network traffic and detect anomalies.

Data Collection: The research will leverage both the **UNSW-NB15** and **KDD CUP 1999** datasets.

Statistical Tools: The model employ statistical analysis to identify distribution patterns in network traffic.

ANOVA: Used to determine if there are statistically significant differences in mean values across different attack categories.

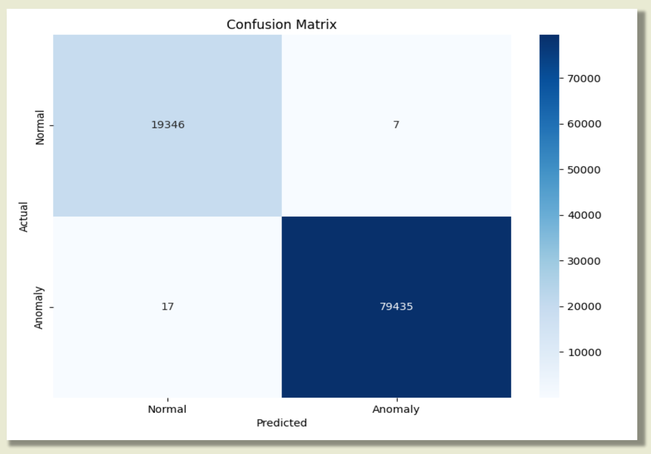
- **Chi-Square Test**: Applied to evaluate the association between categorical variables .
- **Random Forest**: A machine learning algorithm used to classify traffic .
- **Isolation Forest**: A technique for detecting anomalies by isolating outliers faster than normal points.



Confusion Matrix

evaluates the model's performance by comparing predictions to actual labels:

- **True Positives (TP):** Correctly identified anomalies.
- **True Negatives (TN):** Correctly identified normal data.
- **False Positives (FP):** Normal data misclassified as anomalies.
- **False Negatives (FN):** Anomalies misclassified as normal.



Heatmap

The correlation coefficient, often denoted as r, between two variables x and y is computed using the **Pearson correlation formula**:

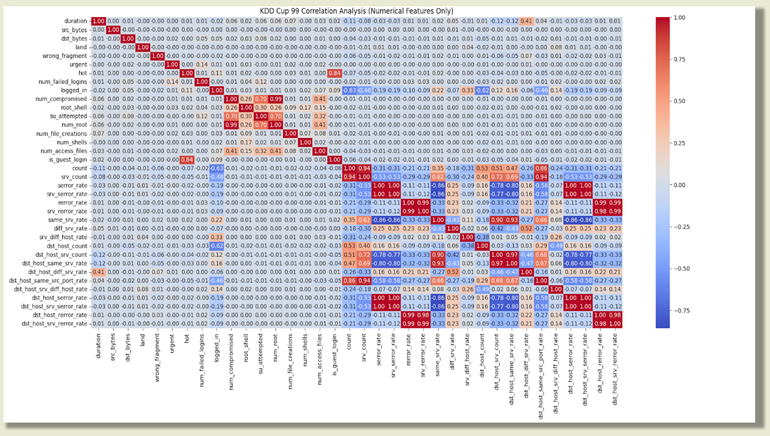
$$r = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

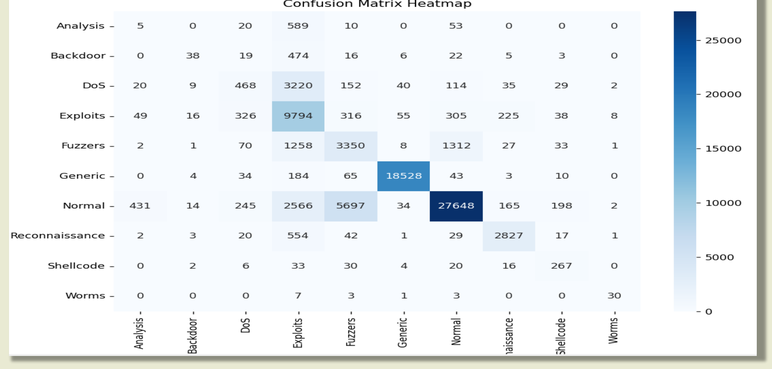
To generate a correlation heatmap:
We compute the pairwise Pearson correlation coefficients for all features in the dataset:

1.For a dataset with m features, compute an $m * m$ C matrix, where:

$$C[i, j] = \frac{\mathrm{Cov}(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}}$$

2. C is symmetric with diagonal elements $C[i, i] = 1 C[i, i] = 1 C[i, i] = 1$ (correlation of a feature with itself).



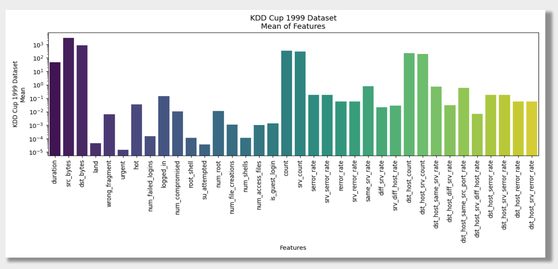


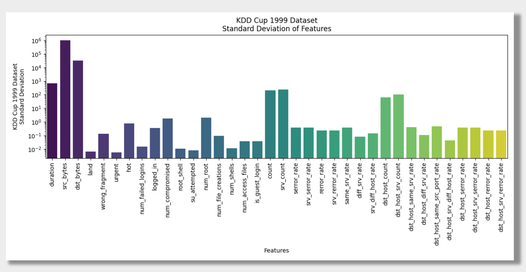
Conclusion

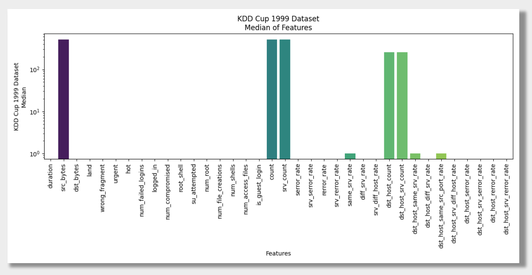
we explored the significance of anomaly detection in network systems as a critical approach to identifying unusual patterns that may indicate security threats, system malfunctions, or performance issues. The study highlights the growing importance of real-time anomaly detection in safeguarding network integrity amidst increasingly complex network environments and evolving cyber threats.
In conclusion, implementing a robust anomaly detection framework that combines accuracy, efficiency, and adaptability is essential for modern networks.

Central Tendency of Features

By training our model on 10% of our dataset we get that the mean of accuracies 85%







Resources:

Team Members:

- Amr Samy Abd Elkarim
- Ali Eldien Alaa Zaki
- Abdullah Ayman Abdelrahman
- Seif Eldien Mohamed Refaat
- Esraa Hassan Ragaa
- Nawal Hossam Mohamed