Cybersecurity Threat Classification Using Machine Learning

Abstract— **This research project develops an intelligent Network Intrusion Detection System (NIDS) using machine learning algorithms to automatically identify different types of cyberattacks. We implemented and compared three machine learning models - Random Forest, Support Vector Machine (SVM), and Neural Network (MLP) - on a synthetically generated network traffic dataset containing normal traffic and three attack types: DDoS, PortScan, and BruteForce. Our experimental results demonstrate that the Neural Network achieved the highest classification accuracy of 96%, followed closely by Random Forest at 95%. The study provides valuable insights into feature importance patterns for different attack types and offers practical recommendations for implementing ML-based security solutions in real-world network environments.**

I. INTRODUCTION

In today's digitally connected world, network security has become paramount as cyber threats grow increasingly sophisticated. Traditional signature-based intrusion detection systems often struggle to keep pace with evolving attack techniques, particularly zero-day exploits. Machine learning offers a promising alternative by learning patterns from network traffic data and detecting anomalies that may indicate malicious activity.

This project investigates the effectiveness of different machine learning approaches for network intrusion detection. We focus on three common attack types that pose significant threats to network infrastructure: Distributed Denial of Service (DDoS) attacks that overwhelm systems with traffic, PortScan attacks that probe networks for vulnerabilities, and BruteForce attacks that attempt unauthorized access through credential stuffing.

II. PROBLEM STATEMENT

Modern networks face several critical challenges in threat detection:

- Existing rule-based systems frequently generate false positives, flagging legitimate traffic as malicious while sometimes missing actual attacks.
- The rapid evolution of attack techniques makes it difficult for traditional systems to detect novel threats without constant rule updates.
- Many current solutions lack the adaptability to recognize subtle patterns in network behavior that may indicate emerging threats.

These limitations create significant security gaps that machine learning approaches can potentially address by learning complex patterns in network traffic and adapting to new threat signatures.

III. OBJECTIVES

The primary objectives of this research were:

1. To develop a machine learning framework capable of accurately classifying different types of network intrusions from traffic data.
2. To compare the effectiveness of three distinct machine learning algorithms (Random Forest, SVM, and Neural Network) in detecting and classifying network attacks.
3. To identify which network traffic features are most indicative of different attack types through feature importance analysis.
4. To establish performance benchmarks for ML-based intrusion detection that can guide future implementations in production environments.

IV. RESEARCH METHODOLOGY

Dataset Description

We generated a synthetic dataset of 10,000 network traffic samples, each characterized by 20 distinct features. The dataset was carefully constructed to simulate real-world network conditions:

- 70% of samples represented normal network traffic
- 10% simulated DDoS attacks
- 10% simulated PortScan attempts
- 10% simulated BruteForce attacks

Attack patterns were injected using specific feature modifications:

- DDoS attacks showed high values in the first 5 features with low variance in the next 5
- PortScan attacks exhibited spikes in middle features (6-10) with uniform low values in features 1-5

- BruteForce attacks displayed discrete, rounded values in the last 5 features

Data Preprocessing

The data underwent comprehensive preprocessing:

1. Missing value handling (though none were present in the synthetic data)
2. Label encoding (converting attack categories to numerical values)
3. Feature scaling using StandardScaler for normalization
4. Feature selection retaining the top 10 most significant features using ANOVA F-test
5. Stratified train-test split (70% training, 30% testing) to maintain class distribution

Machine Learning Models

We implemented three classification algorithms with optimized hyperparameters:

1. **Random Forest**
   - 200 decision trees
   - Maximum depth of 15
   - Minimum samples split of 5
   - Balanced class weights
2. **Support Vector Machine (SVM)**
   - RBF kernel
   - C=1.0
   - Class weight balancing
   - Probability estimates enabled
3. **Neural Network (MLP)**
   - Two hidden layers (64 → 32 neurons)
   - ReLU activation
   - Adam optimizer
   - Early stopping with validation

V. EVALUATION METRICS

We assessed model performance using five key metrics:

1. **Accuracy**: Overall classification correctness
2. **Precision**: Measure of false positives (higher is better)
3. **Recall**: Measure of false negatives (higher is better)
4. **F1-Score**: Harmonic mean of precision and recall

ROC-AUC: Area under the receiver operating characteristic curve these metrics provide a comprehensive view of model performance across different aspects of classification quality. Users can view extracted text and interact with the AI conversational interface by asking questions based on the content.

VII. RESULTS AND DISCUSSIONS

Our experiments yielded the following performance results:

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.95	0.94	0.95	0.94	0.99
SVM	0.93	0.92	0.93	0.92	0.98
Neural Network	0.96	0.95	0.96	0.95	0.99

Key findings from the evaluation:

- The Neural Network achieved the best overall performance across all metrics
- All models showed strong capability in detecting DDoS attacks due to their distinct feature patterns
- BruteForce attacks proved more challenging to detect, with slightly higher false negative rates
- Feature importance analysis revealed that different models weighted features differently, but all identified the first 5 features as most significant for DDoS detection.

Performance Characteristics

- **Neural Network**: Highest accuracy (96%) but requires more computational resources
- **Random Forest**: Nearly as accurate (95%) with faster training times
- **SVM**: Good performance (93%) but slower training compared to others

Practical Considerations

- **Interpretability**: Random Forest offers better feature importance insights
- **Training Speed**: Random Forest trains fastest, Neural Network slowest
- **Resource Requirements**: Neural Network demands most memory and processing power

VII. CONCLUSION

This research demonstrates that machine learning can effectively detect network intrusions, with Neural Networks achieving the highest accuracy (96%) among the tested algorithms. The study confirms that different attack types create distinct patterns in network traffic features that machine learning models can learn to recognize.