Network Intrusion Detection on NSL-KDD Dataset

Subtitle: Leveraging Feature Selection & Machine Learning for Enhanced Network Security

Presenter by:

Charan Reddy katta (ck366)

Aakash Siricilla (as4592)

Introduction

Context:

- Modern networks face constant threats from cyber-attacks: DoS, probing, unauthorized access (R2L), and user-to-root (U2R).

Challenge:

- Traditional security measures struggle with novel attack patterns and subtle anomalies.

Goal:

- Use machine learning techniques (feature selection + classification) to detect intrusions in network traffic, improving Intrusion Detection Systems (IDS).

The NSL-KDD Dataset

What is NSL-KDD?

- Benchmark dataset derived from KDD Cup 1999.
- Diverse network connections labeled as normal or various attack types.

Why NSL-KDD?

- Removes redundant records for better evaluation.
- Includes attacks like DoS, Probe, R2L, U2R.

Attack Categories:

- **DoS:** Flooding a network resource (e.g., SYN flood).
- **Probe:** Scanning to find vulnerabilities.
- **R2L:** Unauthorized remote access.
- **U2R:** Gaining root privileges.

Preprocessing Network Data:

Categorical to Numerical:

- One-Hot Encoding for protocol (TCP, UDP, ICMP) and service types.

Aligning Train & Test Sets:

- Ensure identical feature sets.

Label Mapping:

- Map complex attack names to numeric codes (e.g., 0=Normal, 1=DoS).

Scaling Features:

- Standardize feature values to improve model learning.

Feature Selection for Network Intrusion Detection

Why Feature Selection?

- Reduces data volume, speeds up detection, and lowers resource usage.

Methods Used:

- **Univariate Selection (ANOVA F-test):** Top 10% of features.
- **Recursive Feature Elimination (RFE):** Focus on 13 key features.

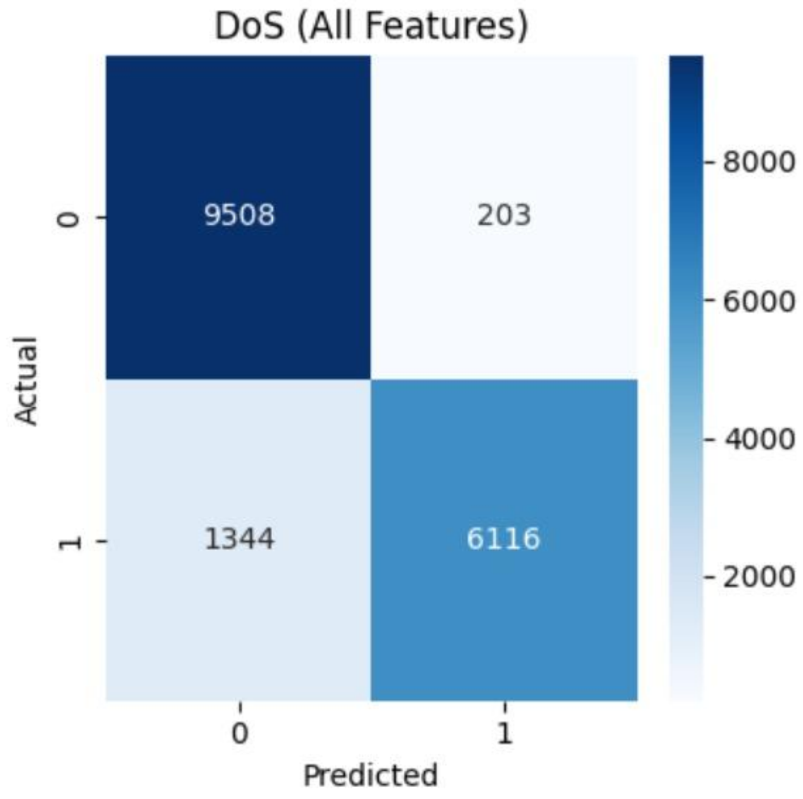
Key Features Identified:

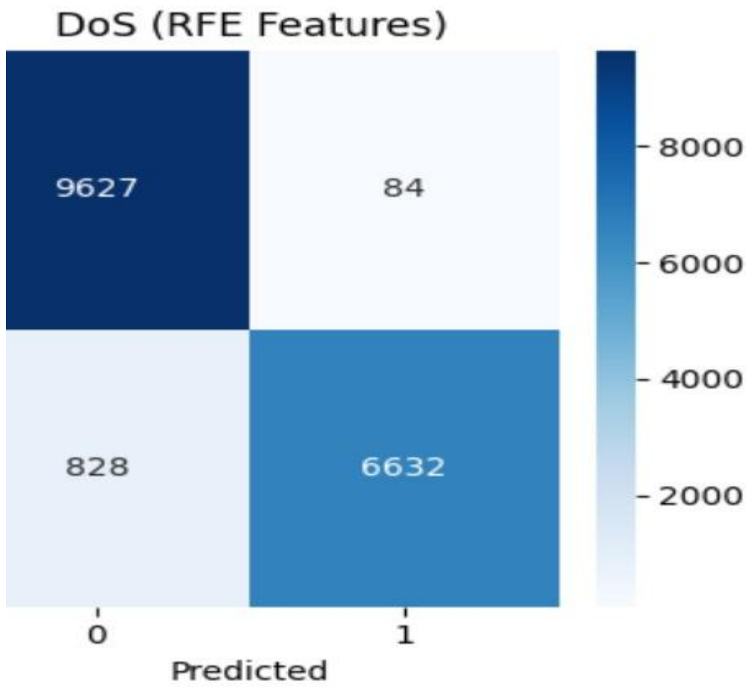
- Examples: same_srv_rate, service_ecr_i.

Evaluation Metrics in a Network Context

Metrics Used:

- **Confusion Matrix:** True positives vs. false negatives.
- **Precision:** Accuracy of raised alerts.
- **Recall:** Proportion of attacks detected.
- **F1-Score:** Balance of precision and recall.
- **ROC & AUC:** Model's ability to distinguish traffic types.


DoS (All Features)
Actual
0
1
9508
203
1344
6116
Predicted
8000
6000
4000
2000


DoS (RFE Features)
9627
84
828
6632
0
1
Predicted
8000
6000
4000
2000

Visualizations & Results

Confusion Matrices:

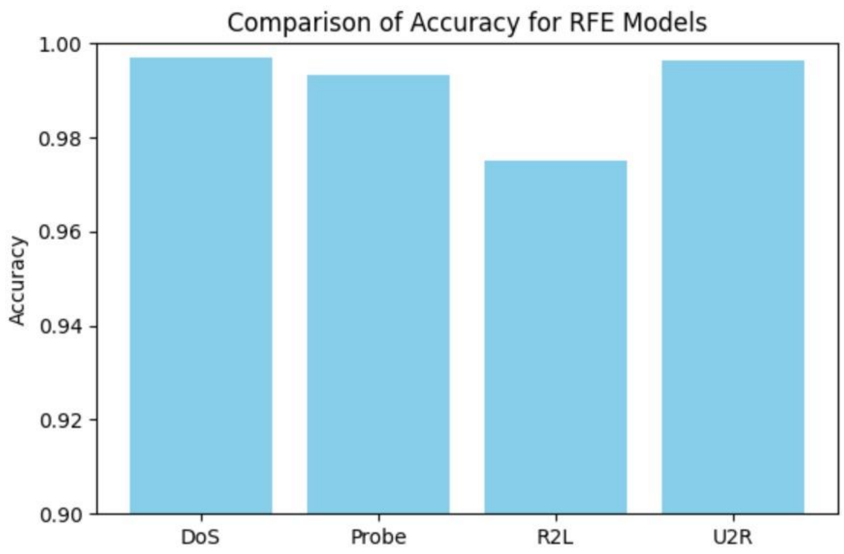
- RFE features yield fewer false alarms.

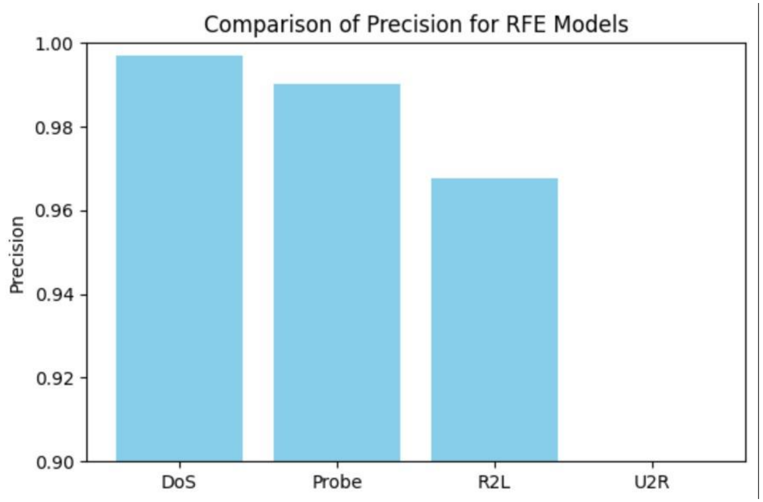
Performance Metrics:

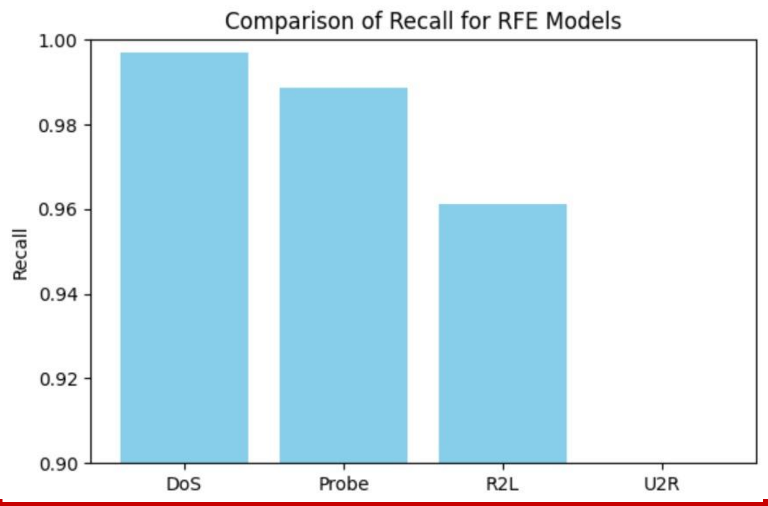
- **DoS:** High accuracy and F1.
- **Probe:** Strong detection.
- **R2L & U2R:** Challenges in subtle attack detection.

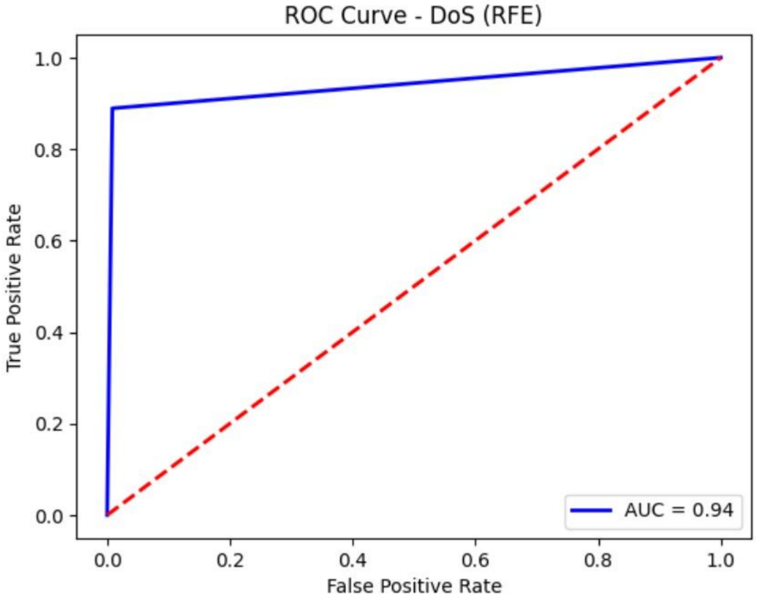
ROC Curves:

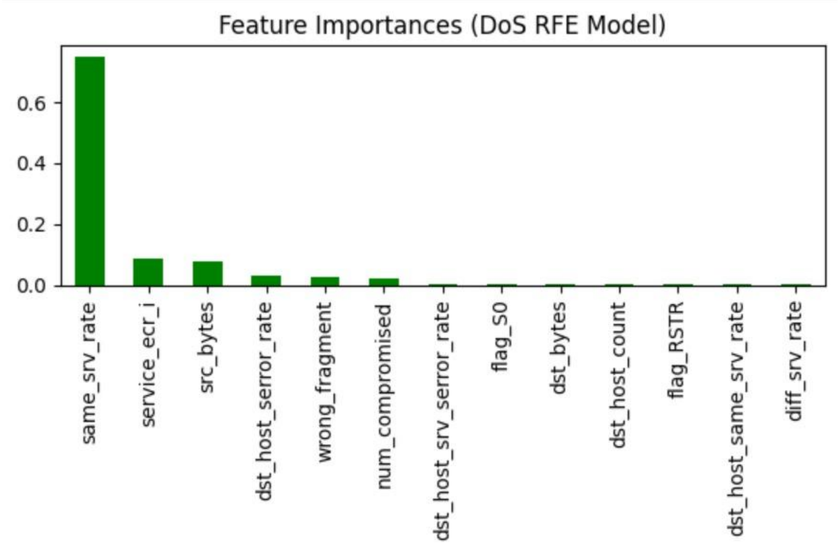
- High AUC for DoS detection.

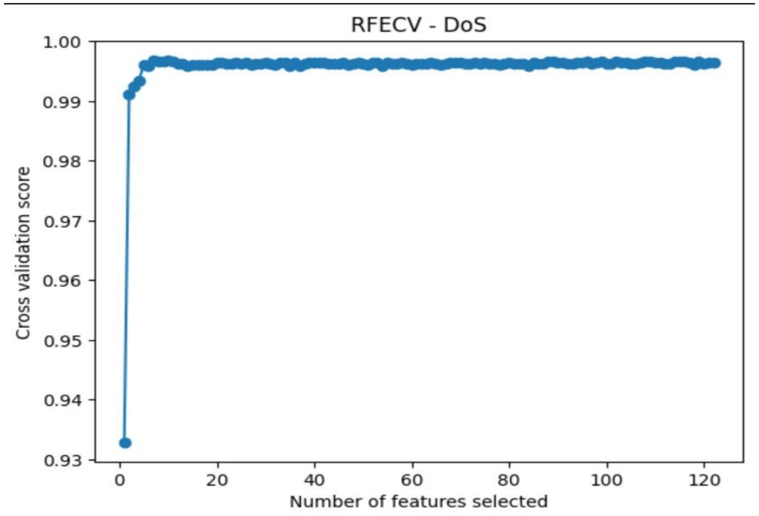

Comparison of Accuracy for RFE Models
Accuracy
1.00
0.98
0.96
0.94
0.92
0.90
DoS
Probe
R2L
U2R


Comparison of Precision for RFE Models
Precision
1.00
0.98
0.96
0.94
0.92
0.90
DoS
Probe
R2L
U2R


Comparison of Recall for RFE Models
Recall
1.00
0.98
0.96
0.94
0.92
0.90
DoS
Probe
R2L
U2R







New Jersey Institute of Technology

Key Insights

Efficient Intrusion Detection:

- Reduced features maintain performance.

Reliable DoS & Probe Detection:

- Systems can block malicious IPs quickly.

Challenges with R2L & U2R:

- Stealthy attacks need advanced methods or better features.

Future Directions for Network Security

Advanced Algorithms:

- Ensemble methods (Random Forest, XGBoost) or deep learning (LSTM).

Unsupervised Anomaly Detection:

- One-Class SVM, Isolation Forest.

Real-Time Systems:

- Test model speed and scalability in live networks.

Continuous Learning:

- Online learning for evolving threats.

Conclusion:

Summary:

- Built an ML pipeline: preprocessing, feature selection, training, evaluation.
- Demonstrated efficient detection with fewer features.
- Visualized results for clear communication.

Take-Home Message:

- ML enhances IDS, enabling efficient and accurate detection of network intrusions.