Towards Smarter Cyber Defense: Anomaly Detection with Autoencoders and LLM-Driven Threat Intelligence

Riley Bruce, Katy Bohanan, Ransom Ward, Prathyu Adari
Computer Science, School of Science and Engineering

Introduction & Background

Network intrusion poses a significant threat to the security and integrity of information systems. Traditional NIDS may struggle with high false-positive rates and fail to detect new, sophisticated attacks. Enhancing detection systems ensures better protection against evolving cyber threats, reduces response times, and improves overall network security, making it crucial for businesses and organizations to safeguard their data and infrastructure.

Our goal is to develop an advanced security system that integrates machine learning (ML) and large language models (LLMs) to enhance real-time threat detection, analysis, and response. This fusion aims to provide a seamless and intelligent security solution.

Datasets

LUFlow NIDS dataset
Encompasses an extensive and diverse number of traffic patterns with information on benign and malicious traffic, as well as outlier behaviour that may or may not be problematic.

Types of network data in LUFlow

- **Avg Input** - Average interpacket time (Timing between packets)
- **Bytes in** - Total number of bytes received from the connection
- **Bytes out** - Total number of bytes sent back to the connection
- **Entropy** - Shannon entropy of packets (randomness measure)
- **Number of packets in** - Total number of packets received
- **Number of packets out** - Total number of packets sent
- **Protocol used** - which network protocol that was used
- **Total entropy** - Overall entropy of the flow of packets
- **Total duration** - Time for completion of the network request
- **Label** - Whether the traffic was benign, malicious, or an outlier

References

Lotfi, S., Modirrousta, M., Shashaani, S., and Aliyari Shoorehdeli, M. "Network Intrusion Detection with Limited Labeled Data Using Self-supervision." arXiv, 2022, https://arxiv.org/pdf/2209.03147.

Chinnasamy, Ramya, et al. "Deep Learning-Driven Methods for Network-Based Intrusion Detection Systems: A Systematic Review." ScienceDirect, Elsevier, https://www.sciencedirect.com/science/article/pii/S2405959525000050.

Xu, Hanxiang, et al. "Large Language Models for Cyber Security: A Systematic Literature Review." arXiv, 2024, https://arxiv.org/pdf/2405.04760.

Methodology

Data & Model Training: Used deep learning techniques to develop models capable of classifying network traffic in real-time. The autoencoder model was trained on the LUFlow dataset to distinguish between normal and anomalous traffic patterns, while the embedding model was trained on a dataset that was balanced to include equal amounts of benign, malicious, and outlier data..

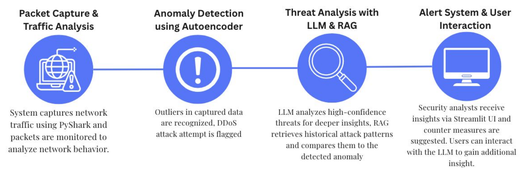
Packet Capture & Traffic Analysis: Used Wireshark and PyShark to capture live network traffic, ensuring accurate, real-time data collection.

Automated Threat Detection: Processed network traffic data through the trained model(s) to predict potential threats in real time. Autoencoder helps identify unknown or novel attack patterns.

LLM-Powered Analysis: Integrates a Large Language Model to interpret detected threats and provide insights and context for threat analysis, as well as help determine further actions to be taken.

Threat Intelligence with RAG: Implemented Retrieval-Augmented Generation (RAG) to reference known attack patterns and intrusion tactics for improved threat context and decision-making.

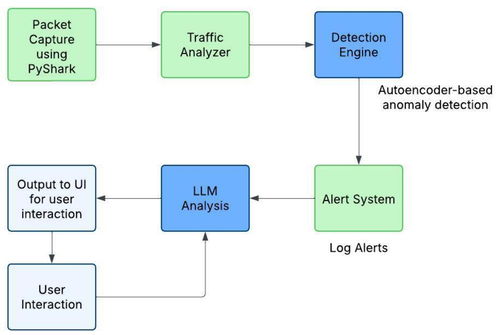
Case Study: DDoS Attack



Normal Traffic vs. Anomalous Traffic

Feature	Normal Traffic	Anomalous Traffic
Source/ Destination	Trusted IPs & Domains	Unknown, spoofed, or blacklisted IPs
Protocols Used	Standard (HTTP, HTTPS, SSH)	Unexpected protocol interactions (DNS Tunneling, ICMP)
Traffic Pattern	Consistent, predictable flow	Sudden spikes, irregular bursts
Packet Size	Expected sizes based on activity	Extremely large or abnormally small payloads
Behavior	Regular usage patterns	Repeated failed logins, port scanning, or DDoS attacks

System Design Flowchart



Conclusion

- Input: Network traffic packets captured in real time
- Processing: Autoencoders detect anomalies, LLM analyzes high-confidence alerts, RAG retrieves historical attack data
- Output: Flagged network threats with contextual analysis, reducing false positives and improving incident response
- Outcome:
  - Successfully identifies known and unknown attack patterns
  - Provides contextual threat insights for better security decisions

Future Work

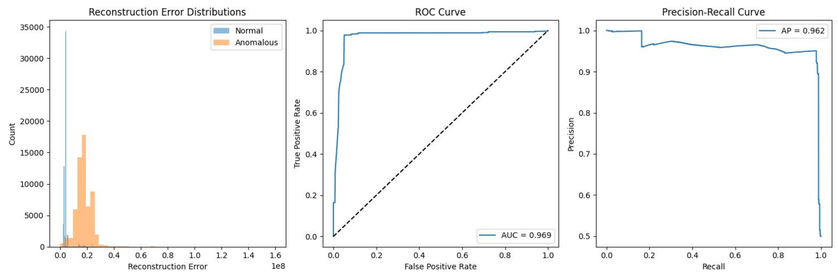
- Enhance Visualization: Develop clearer anomaly reports and interactive dashboards for better insights
- Expand Attack Coverage: Use additional datasets to train the model on a wider variety of network traffic, improving detection accuracy across diverse attack types

Acknowledgements

Dr. Yugyung Lee
5542 Big Data & Analytics course

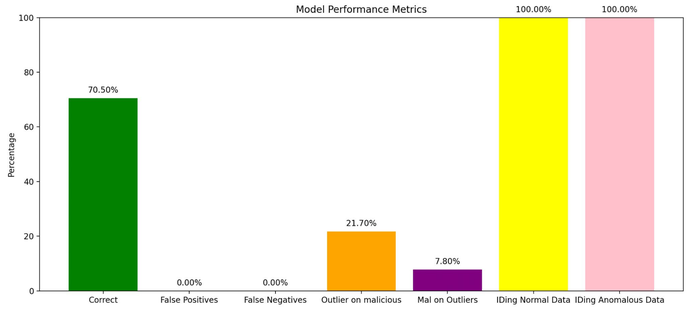
Results & Discussion

Autoencoder Performance Evaluation



- Left: Reconstruction Error Distributions. The distinct separation between normal (blue) and anomalous (orange) reconstruction errors demonstrates the model's ability to identify deviations in network traffic.
- Center: ROC Curve. The near-perfect ROC curve, with an AUC of 0.9969, indicates excellent discriminative power between normal and anomalous network behaviors.
- Right: Precision-Recall Curve. A high Average Precision of 0.962 highlights the model's effectiveness in accurately detecting intrusions with minimal false positives.

Embedding / Feed forward Neural Network Performance



- (This model only got confused on exactly classifying outliers vs malicious)
- This model got about 70% accuracy on correctly identifying all information exactly (Tested on10,000 data points to get accurate results representation)
- But it was able to perfectly separate normal data from anomalous data, meaning it would only flag outliers and malicious data which you typically want a NIDS system to do for further inspection (to be done with the LLM)