Model Research Report – AI-Based Cyber Security Threats Prediction

This document outlines research on machine learning and deep learning models that can be applied to cybersecurity threat prediction. It also discusses the types of data sources (static and real-time) suitable for building AI-powered threat detection and prediction systems.

Model Research

1. **Random Forest Classifier** – Effective for structured security logs, easy interpretability, good baseline model.
2. **Logistic Regression** – Simple, fast to train, effective for binary threat classification.
3. **Support Vector Machine (SVM)** – High accuracy in classifying attack vs. non-attack traffic.
4. **Recurrent Neural Networks (RNN / LSTM)** – Handles sequential data like network flows; useful for anomaly detection.
5. **Autoencoders (Deep Learning)** – Good for detecting unknown or zero-day attacks.
6. **Graph Neural Networks (GNN)** – Useful for modeling network topology and detecting lateral movements.
7. **Hybrid Models (e.g., CNN + LSTM)** – Capture both spatial and temporal patterns in network traffic.

Data Resources

Static Data Sources:
• CICIDS2017 – Realistic labeled intrusion detection dataset with various attack types.
• UNSW-NB15 – Rich modern dataset covering multiple attack categories and normal traffic.
Real-Time Data Sources:
• Zeek (formerly Bro) – Open-source network security monitoring tool that generates real-time network logs.
• Security Onion – Platform for live network traffic capture, intrusion detection, and monitoring.
These sources are ideal for both training models and validating performance in practical scenarios.

Why These Data Sources Are Best

- **CICIDS2017:** Comprehensive coverage of modern cyber-attacks, clean labeling, widely used in research.
- **UNSW-NB15:** Updated dataset, realistic traffic, supports a variety of ML algorithms.
- **Zeek:** Real-time packet inspection and metadata extraction; ideal for streaming ML pipelines.
- **Security Onion:** End-to-end real-time monitoring and threat detection environment.

Conclusion

Selecting the right model and data source is critical in AI-based cyber threat prediction. Combining traditional ML with deep learning methods and using both static and real-time datasets enables more robust and adaptive defense systems.