Cybersecurity: Suspicious Web Threat Interactions

Name: Hardik Singh

Institution: Unified Mentor

Date: July 07, 2025

GitHub Repository

Dataset Link

Objective

The goal of this project is to detect and analyze patterns in web interactions for identifying suspicious or potentially harmful activities. By using deep learning and network-based feature engineering, the system aims to distinguish between legitimate and malicious network behavior.

Dataset Information

- Source: Google Drive
- Shape: 282 rows × 16 columns
- Key Features Include:
  - bytes_in: Bytes received by the server
  - bytes_out: Bytes sent from the server
  - creation_time: Record creation time
  - end_time: End time of the connection
  - src_ip: Source IP address
  - src_ip_country_code: Country of source IP
  - protocol: Protocol used
  - response.code: HTTP response code
  - dst_port: Server destination port
  - dst_ip: Destination IP
  - rule_names, observation_name, source.meta, source.name, detection_types

Workflow & Methodology

1. Data Cleaning:
   - Formatted timestamps, removed inconsistencies

2. Exploratory Data Analysis (EDA):

- Visualized protocol distribution, traffic anomalies, response codes

3. Feature Engineering:

- Derived threat intensity and protocol-based risk indicators

4. Encoding:

- Encoded categorical fields like detection_types, protocol

5. Deep Learning Modeling:

- Built a binary classifier using TensorFlow to predict suspicious activities

Model Used

- Deep Learning (TensorFlow)

- Optimized with categorical cross-entropy and early stopping

Tools & Technologies Used

- Python, Jupyter Notebook

- Libraries: Pandas, NumPy, Matplotlib, Seaborn, Sklearn, TensorFlow

Challenges Faced

- Small dataset size required careful validation to avoid overfitting

- Complex feature interactions between protocols and rule-based flags

Conclusion

Deep learning models can effectively detect patterns in network interaction data. The project successfully demonstrated the viability of automated cybersecurity monitoring using structured logs and rule-based labeling.

Future Improvements

- Collect larger, real-time datasets for more robust training

- Integrate attention-based models for contextual threat detection

- Deploy as an API for enterprise use