Cybersecurity Threat Analysis

22nd October 2024

PROJECT OVERVIEW:

This project aims to analyse a cybersecurity dataset to understand better the types of attacks, the frequency of occurrences, targets, and the effectiveness of preventative measures. The analysis will help detect cyber-attack patterns, assess their damage, and evaluate how well different cybersecurity protocols have mitigated these threats.

OBJECTIVES:

1. **Distribution of Attack Types:**
    - **Objective:** Categorize and analyze the distribution of various attack types (e.g., DDoS, phishing, malware) to understand which are most preventable.
    - **Approach:** Group the attack types from the dataset and visualize the distribution to identify the most common types.
2. **Attack Frequency (Time Series Analysis):**
    - **Objective:** Perform a time series analysis on the occurrence of attacks to reveal trends, spikes, and periods with higher attack frequency.
    - **Approach:** Use the timestamp data to create a time series and observe the frequency and trends of attacks over time, identifying peak periods.
3. **Target Analysis (Region and Organization):**
    - Objective: Analyze the geo-location and user information to determine which regions, organizations, or industries are the most targeted.
    - Approach: Use geographic and organisational information to break down attacks by location and industry, helping to identify high-risk areas.
4. **Damage Assess:**
    - Objective: Assess the financial or operational damage caused by attacks, if available in the dataset (via anomaly scores, financial indicators, or other damage matrics).
    - Approach: Compare the damage across different attack types and regions to quantify the impact of each type of cyber attack.

5. **Preventive Measures (Effectiveness of Protocols):**
    - Objective: Evaluate the effectiveness of different cybersecurity measures by examining responses such as blocking or logging attacks.
    - Approach: Analyze the action taken (e.g., blocked, logged) and protocol performance metrics to measure how well different security systems handled the attacks.

DATASET DESCRIPTION:

The dataset contains 40,000 entries with 25 columns, many capturing various technical details related to cyber-attacks.

1. Timestamp: The date and time when the attack occurred.
2. Source IP Address: The IP address of the originator of the attack.
3. Destination IP Address: The IP address of the target system.
4. Source Port: The port number used by the source system.
5. Destination Port: The port number targeted on the destination system.
6. Protocol: The communication protocol used (e.g., TCP, UDP, ICMP).
7. Packet Length: Size of the packet in bytes.
8. Packet Type: Type of packet (e.g., Data, Control).
9. Traffic Type: Type of traffic (e.g., HTTP, DNS).
10. Payload Date: Contents of the packets in bytes.
11. Malware Indicators: Indicators of potential malware in the packet.
12. Anomaly Scores: A score indicating how unusual the network activity is.
13. Alerts/Warning: Alerts or warnings triggered by the attack.
14. Attack Type: The type of attack (e.g., DDoS, Malware)
15. Attack Signature: A unique identifier for the attack pattern.
16. Action Taken: Action taken by the system (e.g., Bocked, Logged).
17. Severity Level: Severity of the attack (e.g., Low, Medium, High).
18. User Information: User associated with the attack or the target.
19. Device Information: Details about the device involved.
20. Network Segment: The network segment where the attack occurred.
21. Geo-location Date: Location of the source IP (city and region).
22. Proxy Information: Details of the proxy used in the attack (if any).
23. Firewall Logs: Logs related to firewall activity during the attack.
24. IDP/IPS Alerts: Alerts generated by the Intrusion Detection/Prevention System.
25. Log Source: The system that logged the events (e.g., Server, Firewall).

This dataset contains information that can be used for a detailed analysis of cybersecurity threats, attack patterns, and the effectiveness of responses.

PROJECT PHASES:

Phase 1: Data Exploration & Cleaning

- Task: Inspect the dataset to identify missing or inconsistent data.
- Deliverable: Cleaned and well-structured dataset ready for analysis.

Phase 2: Distribution of Attack Types

- Task: Categorize the attack types and visualise the distribution.
- Deliverable: Bar/column charts showing the most frequent attack types (e.g., DDoS, Malware, Intrusion).

Phase 3: Attack Frequency Analysis

- Task: Perform a time series analysis to visualize attack frequency over time.
- Deliverable: Line graph or heatmaps showing periods of high or low attack frequency.

Phase 4: Target Analysis by Region and Organization

- Task: Analyze the geographic distribution of attacks across regions and industries.
- Deliverable: Geographical heat maps or charts that identify high-risk regions and organizations.

Phase 5: Damage Assessment

- Task: Assess the impact or damage of attacks across using available data (e.g., anomaly scores, financial losses).
- Deliverable: Comparative tables or charts measuring the damage caused by different attack types.

Phase 6: Effective of Preventive Measures

- Task: Analyze the effectiveness of various cybersecurity protocols.
- Deliverable: Performance metrics or success rates for different measures like blocking, logging, and firewall responses.

Phase 7: Reporting and Visualization

- Task: Summarize key findings and generate visualizations.

- Deliverable: A final report with key insights, charts, and recommendations for improving cybersecurity defences.

TOOLS AND TECHNOLOGIES:

- **Programming Languages:** Python (for analysis and visualisations).
- **Libraries:** Pandas, Matplotlib/Seaborn, Plotly (for visualisations), Scikit-learn (if any machine learning is needed for anomaly detection).
- **Data Sources:** Dataset containing cybersecurity events, including attack types, timestamps, geo-location data, and response actions.

EXPECTED OUTCOME:

The project will result in a comprehensive analysis of cybersecurity threats, highlighting the most common attacks, vulnerable regions and organisations, their impact, and the effectiveness of existing defences. This will aid stakeholders in making data-driven decisions to improve cybersecurity strategies.