# ACENET
## Microcredential in Advanced Computing ISP Report

**Project title: Cyber Attack Analysis**

**Participant name:** Fatema Alasbahi

**Date:** July 31, 2024

**Abstract:**

This project aims to investigate cyber-attack patterns and trends by analyzing a dataset of cybersecurity incidents. The primary objective is to identify the most prevalent attack types, target devices, temporal patterns, and leverage this information to inform effective security strategies.

### 1. Introduction

Cybersecurity has emerged as a critical concern due to the increasing frequency of cyber-attacks. This study focuses on understanding the landscape of cyber threats by analyzing a dataset containing details of various security incidents.

Our aim is to identify the most common attack types and targeted devices, along with uncovering temporal patterns in attack occurrences.

### 2. Background

Cybersecurity threats pose significant challenges to individuals and organizations.

This study concentrates on three primary attack types: Distributed Denial of Service (DDoS), malware, and intrusion attacks. DDoS attacks overwhelm systems with traffic, causing service disruptions and financial losses. Malware, including viruses, worms, and ransomware, can compromise system integrity, steal data, and disrupt operations. Intrusion attacks aim to gain unauthorized access to computer systems, leading to

data breaches and espionage. Understanding the characteristics and impact of these attack types is important for creating effective defense strategies.

## 3. Analysis

### 3.1 Data Description:

The data for this analysis was sourced from Kaggle, with over than 40,000 rows and 25 columns.

This dataset contains information on various security incidents, including attack types, devices information, timestamps, severity level and other relevant details.

### 3.2 Data Preprocessing:

- **Missing Values:** I assessed the data for missing values in certain columns and employed appropriate techniques to address them by replacing the missing values in columns like "Alerts/Warnings", "Malware Indicators", "Proxy Information", "Firewall Logs", and "IDS/IPS Alerts" with relevant placeholders like "No Alert Triggered", "No Detection", "No Proxy", "No Data", and "No Data" respectively.

- **Device Information:**

  The "Device Information" column underwent processing to extract the primary device type used to access the system.

  This was achieved by splitting the column based on "/" and storing the first element as the "Browser" in a new column.

  Regular expressions (RE) were then employed to identify common operating systems and devices (Windows, Linux, Android, iPad, iPod, iPhone, Macintosh) within the "Browser" column and a new column named "Device/OS" was created to store the extracted device or operating system.

  Finally, the "Device Information" column was dropped.

### 3.3 Analysis Techniques

#### 3.3.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis was performed to understand the basic characteristics of the dataset and uncover patterns, anomalies, and relationships within the data.

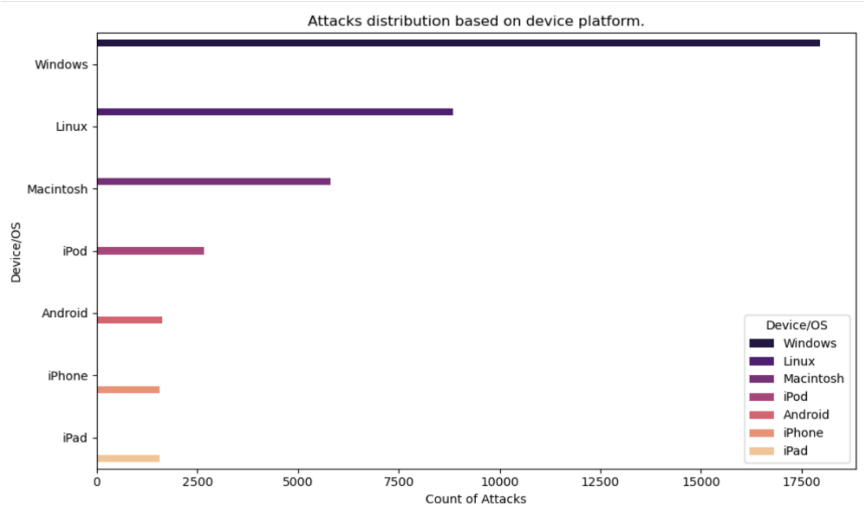
### 3.3.2 Time-Series Analysis

Time-series analysis was performed to identify temporal patterns in the dataset and it was essential for understanding how the frequency and nature of cyber-attacks changed over time.

## 4. Results

### 4.1 Device/OS Distribution:

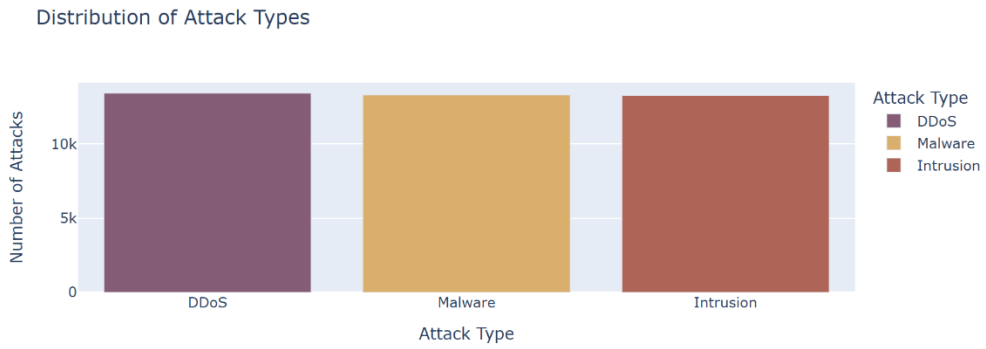
- Analysing the "Device/OS" column revealed Windows as the most prevalent target device, followed by Linux.




### 4.2 Attack Type Distribution:

- By counting occurrences within the "Attack Type" column, we identified Distributed denial of service (DDoS) attacks as the most frequent

threat,followed by malware then intrusion attempts



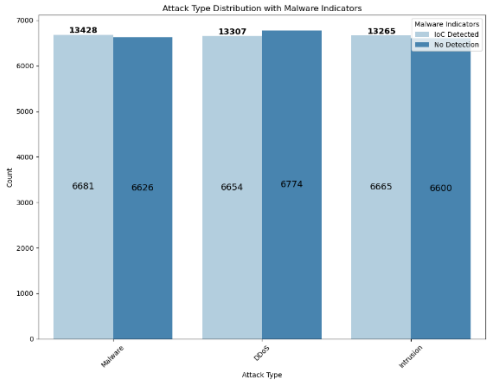

### 4.3 Visualization of Attack Types:

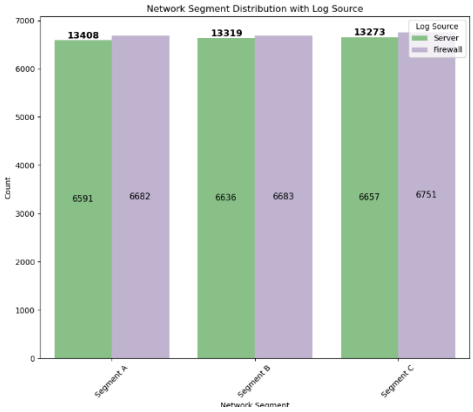
- A bar chart was created to visualize the distribution of attack types and the finding were that DDoS attacks as the dominant threat.

### 4.4 Attack Type and Additional Features:

- Interactive bar charts and grouped bar charts using various libraries (Seaborn, Plotly) to explore correlations between attack types and other features like:
    - Malware Indicators
    - Alerts/Warnings
    - Attack Signature
    - Action Taken
    - Severity Level

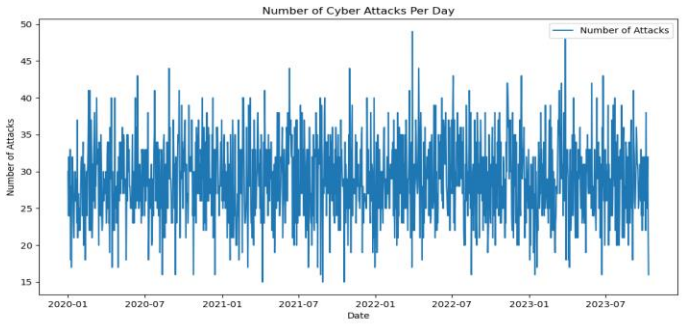
- Log Source (in relation to Network Segment)




## 4.5 Temporal Analysis:

- The "Timestamp" column was converted to a datetime format to extract valuable time-based features like year, month, day, and hour.
- We explored attack occurrences over time by:
  - Analyzing the number of attacks per day.
  - Investigating attacks by hour of the day.
  - Examining attacks by day of the week.
  - Creating a heatmap to visualize attack trends across months in the years 2020 to 2023.




**<u>*NOTE: all the plots are available in the GitHub Repository</u>**

## 5. Discussion

The analysis provided important insights about cyber-attack patterns.

We identified prevalent attack types, target devices, and temporal patterns by using Exploratory Data Analysis (EDA) and Time-Series Analysis, and found that DDoS attacks were the most common.

The dataset's selection and size were a limitation because the small dataset from Kaggle limited the use of the HPC capabilities and restricting the analysis depth.

While the dataset might not cover all types of cybersecurity threats, future research could benefit from larger and more varied datasets.

In summary, while the study provided valuable insights, addressing and focusing on these limitations could improve future research and despite these challenges, the findings emphasize the need for adaptive and proactive cybersecurity strategies.

## Conclusion

This study has provided valuable insights into the cyber-attack landscape.

The findings underscore the dominance of DDoS attacks and the need to focus on securing Windows and Linux systems.

Temporal patterns indicate the importance of continuous monitoring and adaptation. Future research should explore advanced analytics and machine learning techniques to enhance cyber defense capabilities.

## Supplementary Materials

- GitHub Repository: [fatemaalasbahi/Cyber_attack_analysis (github.com)]