LARGE-SCALE ATTACKS DATA ANALYSIS IN IOT ENVIRONMENTS

CICIoT2023 Dataset

Dr. Jongwook Woo

Group 2

Neelam Patidar - Sally Zreiqat - Sirisha Mahesh

Agenda

- Project Overview
- Dataset Details
- Cluster Specification
- Project Workflow
- Analysis & Visualization
- Challenges / Solutions
- Conclusion

Project Overview

Project Objectives:

- ✓ **Attack Analysis**: In-depth analysis of 33 distinct attacks categorized into seven types: DDoS, DoS, Recon, Web-based, Brute Force, Spoofing, and Mirai.

Key Analysis Findings:

- ✓ **Incident Analysis**: Categorized and quantified incidents by attack type and frequency.
- ✓ **Duration Metrics**: Evaluated the average, minimum, and maximum durations of each attack type.
- ✓ **Traffic Metrics**: Analyzed average and range of packet sending and receiving rates.
- ✓ **Protocol Impact**: Assessed which protocols are most affected by attacks, focusing on frequency and type.
- ✓ **Targeted Attacks**: Identified the most frequently targeted attack types, including analysis of flow duration and packet size.
- ✓ **Flag Analysis in Network Traffic**: Focused on different flags (SYN, FIN, ACK) associated with various attack types in the IoT dataset, providing insights into attack signatures and behaviors.

Dataset Details

Some Insight on Dataset:

- ➢ Research presents a detailed IoT attack dataset to advance security analytics in real IoT environments.
- ➢ It details 33 attacks distributed across seven categories: DDoS, DoS, Recon, Web-based, Brute Force, Spoofing, and Mirai.
- ➢ Attacks are conducted by malicious IoT devices targeting other devices in the network.

- ✓ **Name:** CICIoT2023 Dataset
- ✓ **Size:** 2 GB
- ✓ **No. of Files:** 30
- ✓ **Format:** CSV

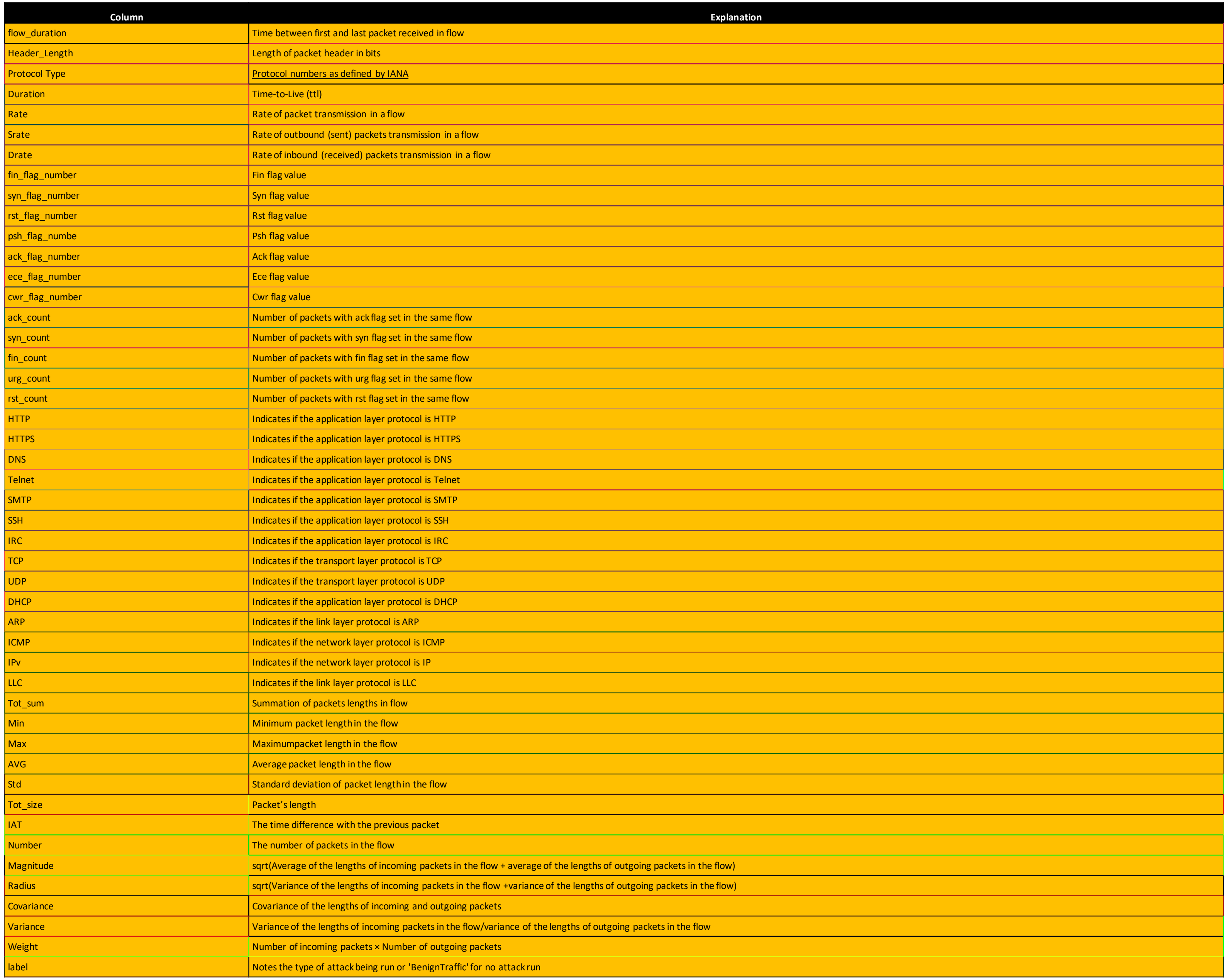
Dataset URL:
https://www.kaggle.com/datasets/madhavmalhotra/unb-cic-iot-dataset

GitHub URL:
NeelamPatidar97/CIS-5200 (github.com)

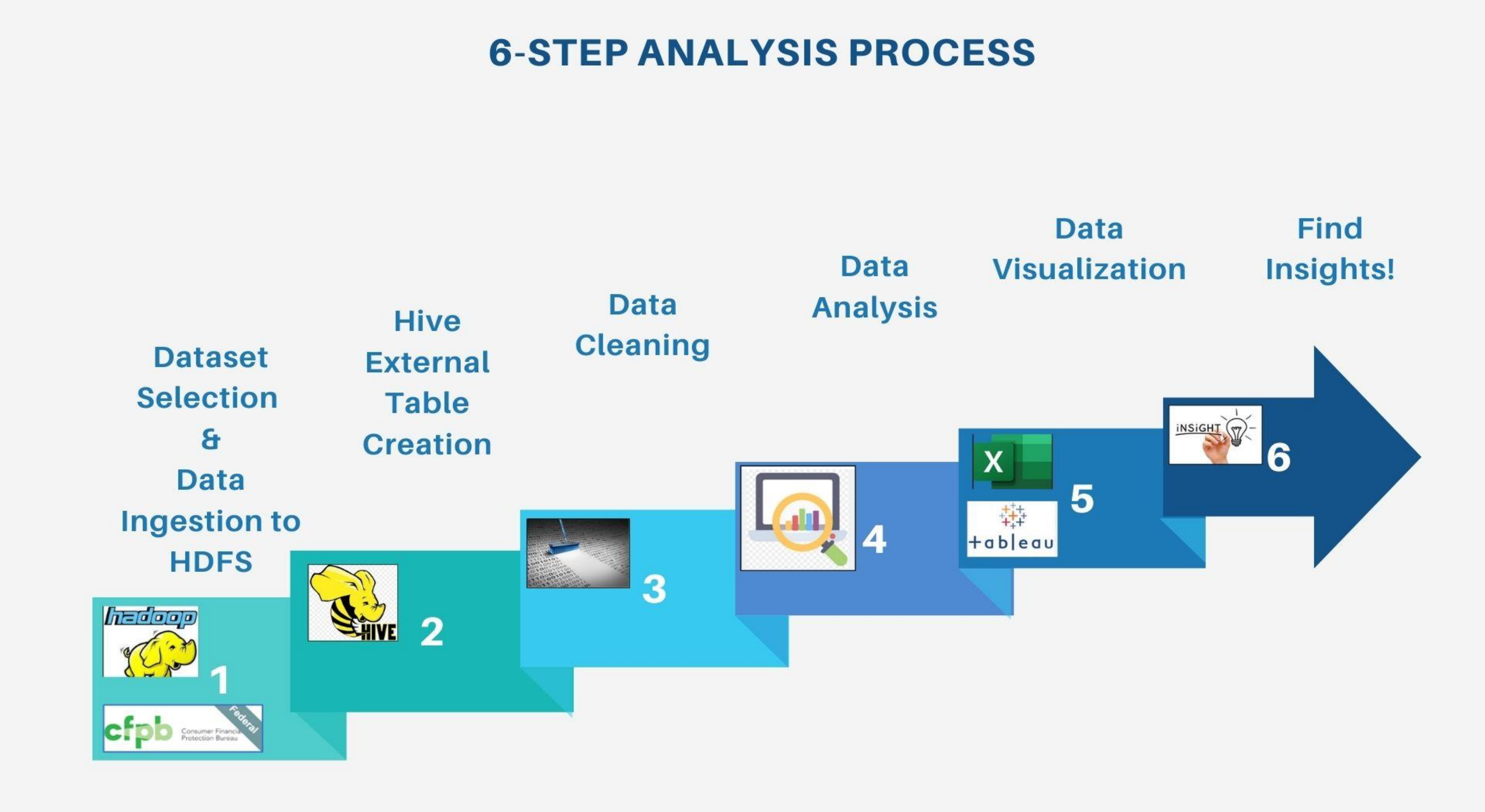
Hadoop Cluster Specifications

- ✓CLUSTER VERSION: Hadoop 3.3.3
- ✓CLUSTER NODES: 5 (2 master nodes & 3 data nodes)
- ✓MEMORY SIZE: Memory Used – 142.45 GB, Memory Remaining – 661.25 GB
- ✓CPU SPEED: 1995.312 MHz

DATASET COLUMNS



Column	Explanation
flow_duration	Time between first and last packet received in flow
Header_Length	Length of packet header in bits
Protocol Type	Protocol numbers as defined by IANA
Duration	Time-to-Live (ttl)
Rate	Rate of packet transmission in a flow
Srate	Rate of outbound (sent) packets transmission in a flow
Drate	Rate of inbound (received) packets transmission in a flow
fin_flag_number	Fin flag value
syn_flag_number	Syn flag value
rst_flag_number	Rst flag value
psh_flag_numbe	Psh flag value
ack_flag_number	Ack flag value
ece_flag_number	Ece flag value
cwr_flag_number	Cwr flag value
ack_count	Number of packets with ack flag set in the same flow
syn_count	Number of packets with syn flag set in the same flow
fin_count	Number of packets with fin flag set in the same flow
urg_count	Number of packets with urg flag set in the same flow
rst_count	Number of packets with rst flag set in the same flow
HTTP	Indicates if the application layer protocol is HTTP
HTTPS	Indicates if the application layer protocol is HTTPS
DNS	Indicates if the application layer protocol is DNS
Telnet	Indicates if the application layer protocol is Telnet
SMTP	Indicates if the application layer protocol is SMTP
SSH	Indicates if the application layer protocol is SSH
IRC	Indicates if the application layer protocol is IRC
TCP	Indicates if the transport layer protocol is TCP
UDP	Indicates if the transport layer protocol is UDP
DHCP	Indicates if the application layer protocol is DHCP
ARP	Indicates if the link layer protocol is ARP
ICMP	Indicates if the network layer protocol is ICMP
IPv	Indicates if the network layer protocol is IP
LLC	Indicates if the link layer protocol is LLC
Tot_sum	Summation of packets lengths in flow
Min	Minimum packet length in the flow
Max	Maximumpacket length in the flow
AVG	Average packet length in the flow
Std	Standard deviation of packet length in the flow
Tot_size	Packet's length
IAT	The time difference with the previous packet
Number	The number of packets in the flow
Magnitude	sqrt(Average of the lengths of incoming packets in the flow + average of the lengths of outgoing packets in the flow)
Radius	sqrt(Variance of the lengths of incoming packets in the flow +variance of the lengths of outgoing packets in the flow)
Covariance	Covariance of the lengths of incoming and outgoing packets
Variance	Variance of the lengths of incoming packets in the flow/variance of the lengths of outgoing packets in the flow
Weight	Number of incoming packets × Number of outgoing packets
label	Notes the type of attack being run or 'BenignTraffic' for no attack run



6-STEP ANALYSIS PROCESS

1 Dataset Selection & Data Ingestion to HDFS

2 Hive External Table Creation

3 Data Cleaning

4 Data Analysis

5 Data Visualization

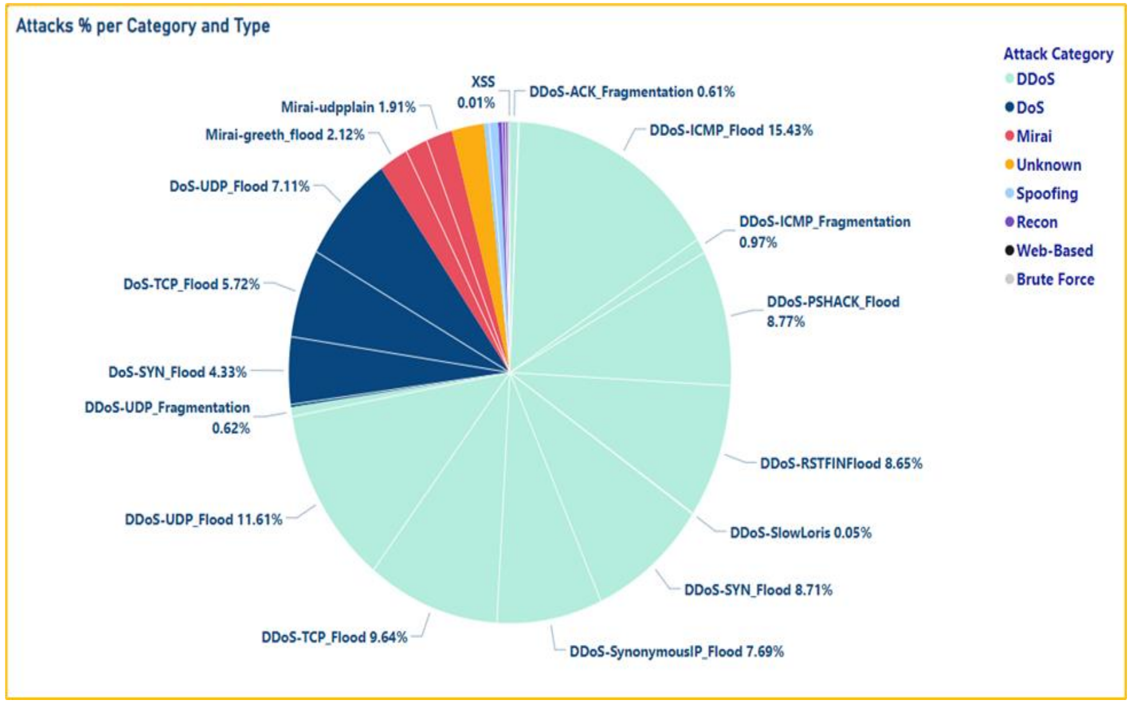
6 Find Insights!

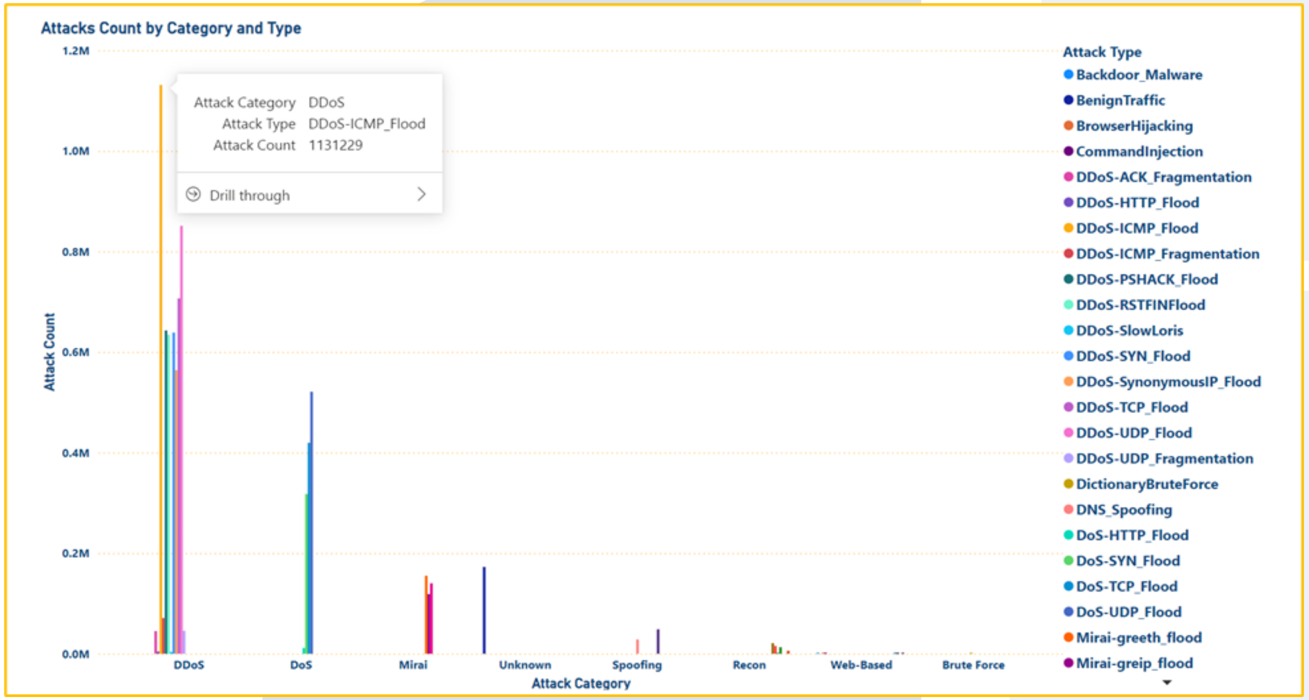
ANALYSIS
&
VISUALISATION

Chart 1: Distribution of Targeted Attacks

Percentage Distribution vs Count of Targeted Attacks by Category and Type

Data Utilized: Attack types, categories and attacks count





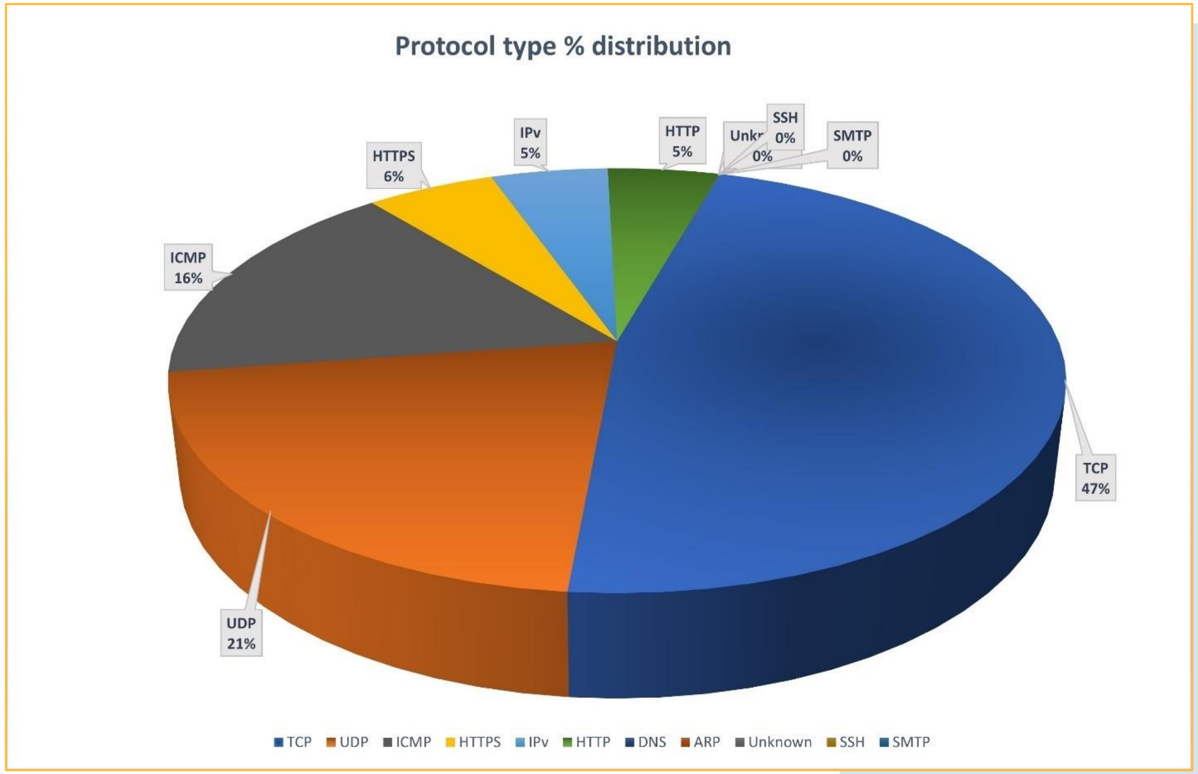
Attack Category DDoS
Attack Type DDoS-ICMP_Flood
Attack Count 1131229

- ✓ **Primary Observation:** DDoS attacks were the predominant form of targeted attacks on various IoT devices. Among these, the DDoS-ICMP Flood was the most frequent, comprising 72.75% of all attacks observed.

- ✓ **Less Frequent Attacks:**
  Web-based Uploading Attacks were the least common, representing less than 0.01% of the attacks.

Chart 2: Percentage of Target Attacks per Protocol

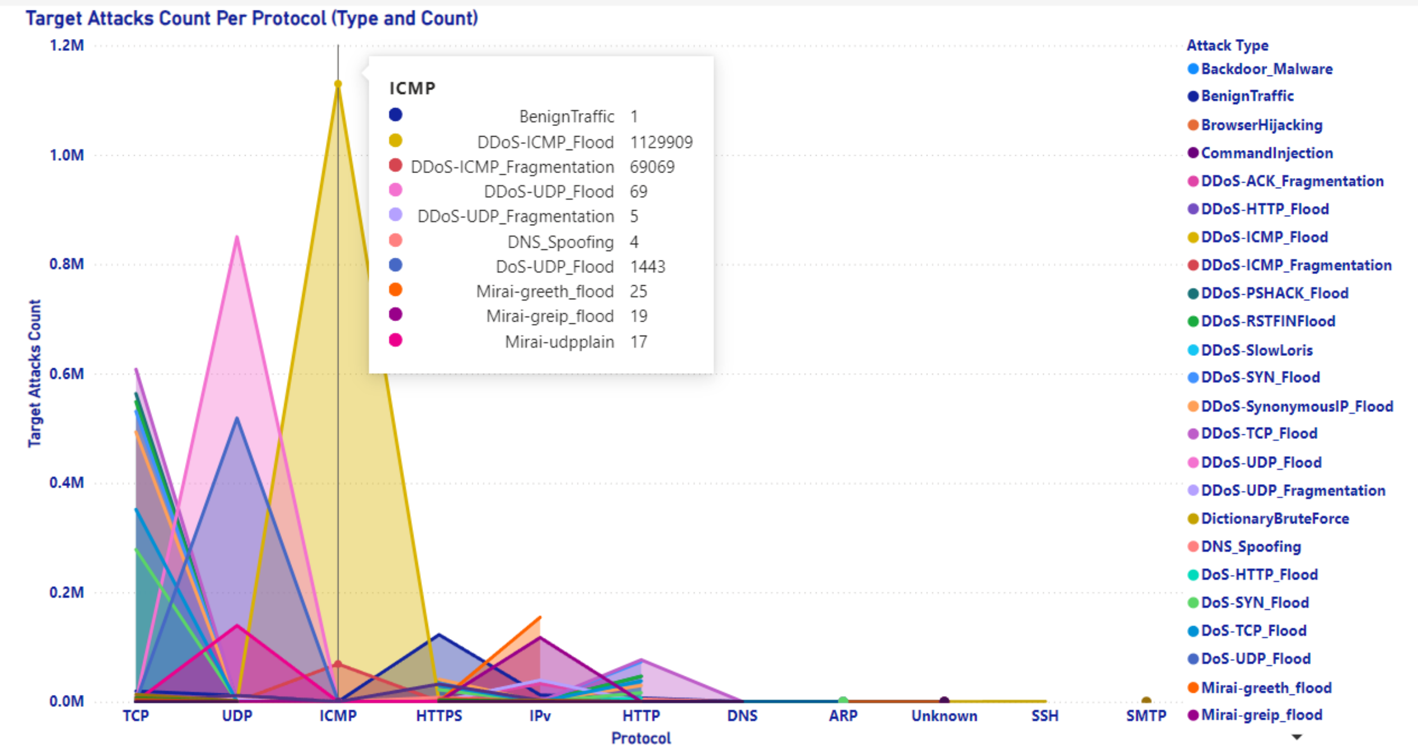
Percentage Distribution of Targeted Attacks by Network Protocol
Data Utilized: Network protocol and count of attacks



- ✓ **Primary Observation:** TCP emerged as the most targeted network protocol, accounting for 47% of all network attacks (vulnerable protocol, requiring extensive security measures and attention).
- ✓ **Less Frequent Attacks:** SSH and SMTP protocols were among the less frequently targeted, indicating lower occurrences compared to TCP (most secure in the IoT ecosystem). Additionally, Web-based Uploading Attacks were particularly rare, representing less than 0.01% of the attacks.

Chart 3: Attack Types Count/Distribution per Protocol

Distribution of Targeted Attacks (by count) per Network Protocol
Data Utilized: Attack type, protocol and count of attacks



ICMP

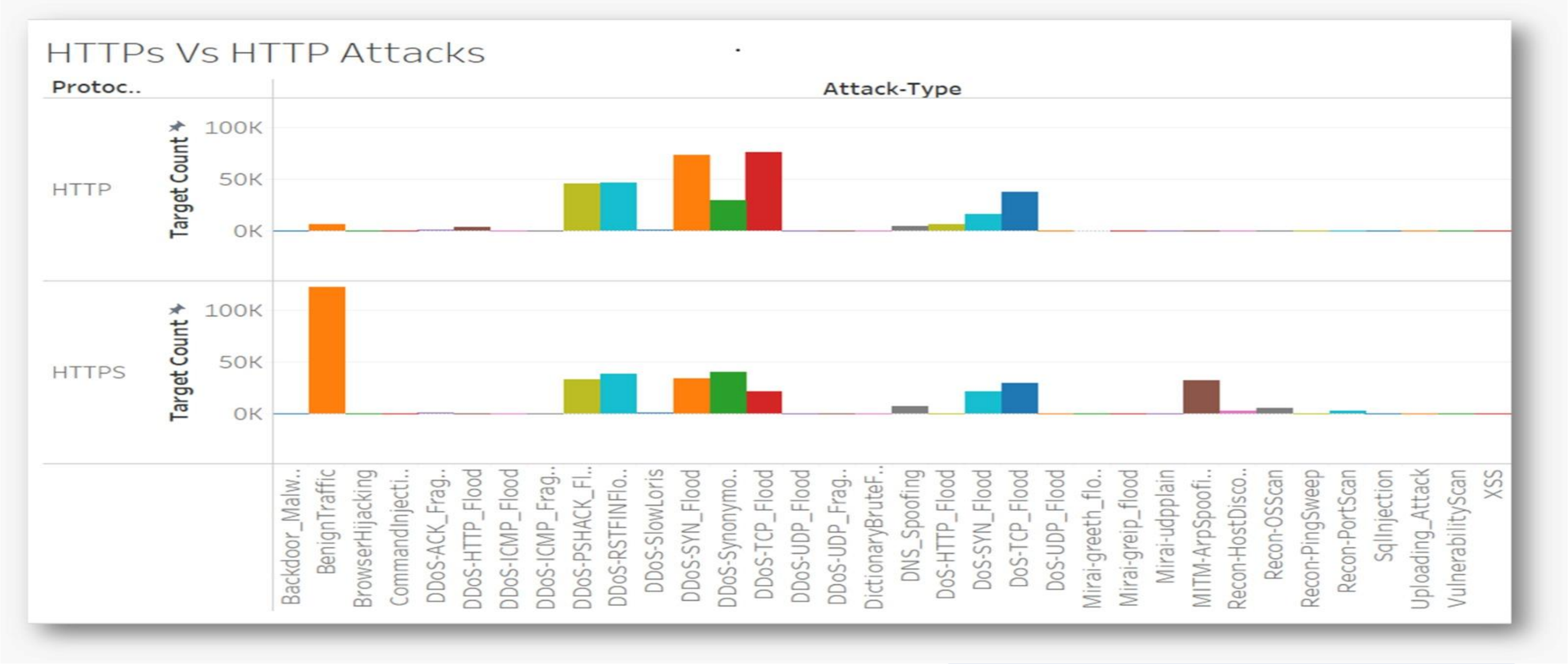
BenignTraffic	1
DDoS-ICMP_Flood	1129909
DDoS-ICMP_Fragmentation	69069
DDoS-UDP_Flood	69
DDoS-UDP_Fragmentation	5
DNS_Spoofing	4
DoS-UDP_Flood	1443
Mirai-greeth_flood	25
Mirai-greip_flood	19
Mirai-udpplain	17

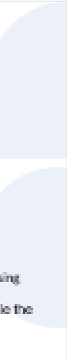
- ✓ **Primary Observation:** ICMP saw a significantly high number of attacks, with a total of 1,129,909 for the type DDoS-ICMP Flood. Meanwhile, TCP was the most frequently targeted network overall, with four attack types each recording occurrences exceeding 500,000.
- ✓ **Less Targeted Networks:** SSH and SMTP were among the least targeted protocols, with total attacks recorded at 328 and 1, respectively.

Chart 4: Comparison between HTTP and HTTPs attacks

Distribution of Targeted Attacks by Category and Type

Data Utilized: Attack type, network protocol and count of attacks

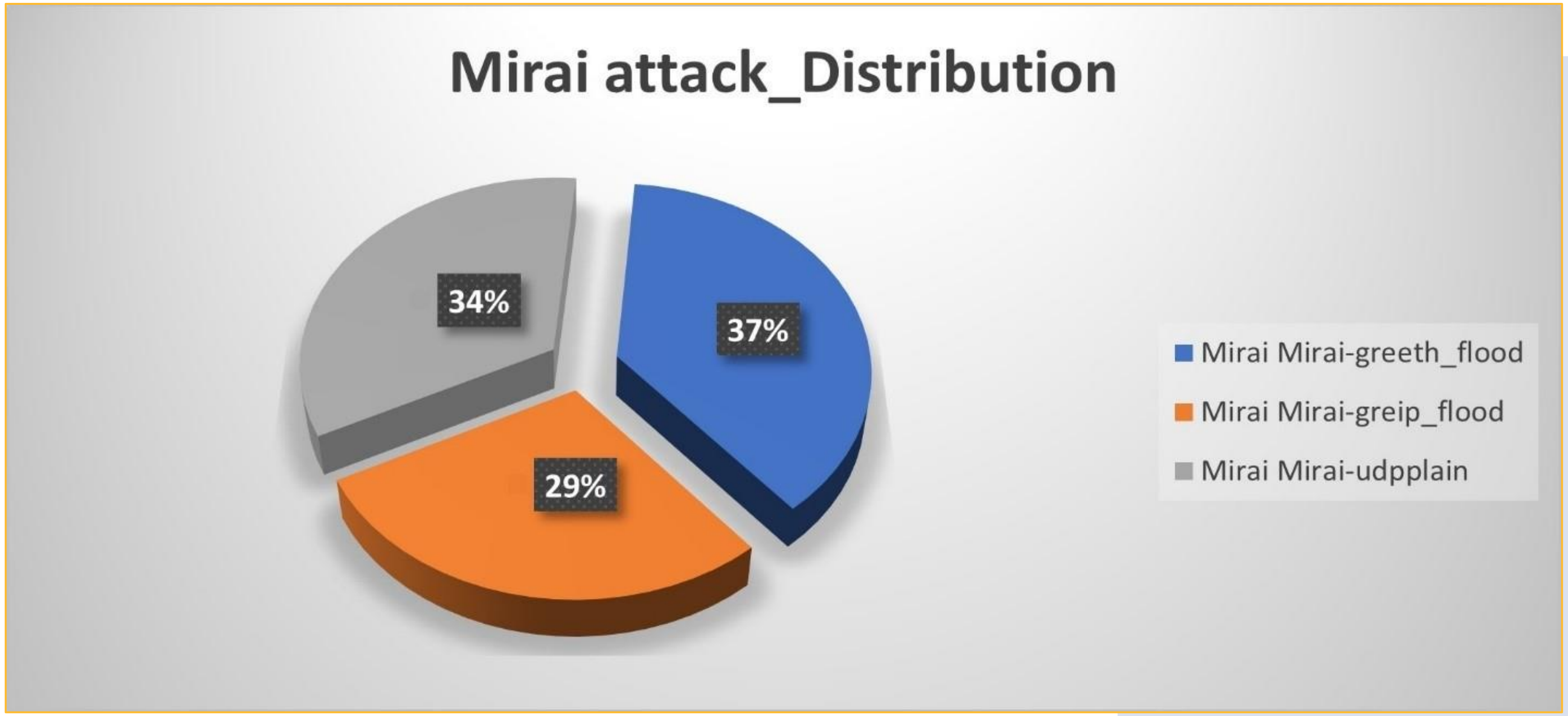




- ✓ **Primary Observation:** Major primary attack we can see in Benign Traffic in HTTPS which is crossing 100k target count while in HTTP the same Benign Traffic is found less than 10k attacks
- ✓ **Less Frequent Attacks:** DDoSFlood_attacks where less number of attacks in HTTPS while the same attacks are found nearly 30K in HTTP.

Chart 5: Mirai Attack Distribution

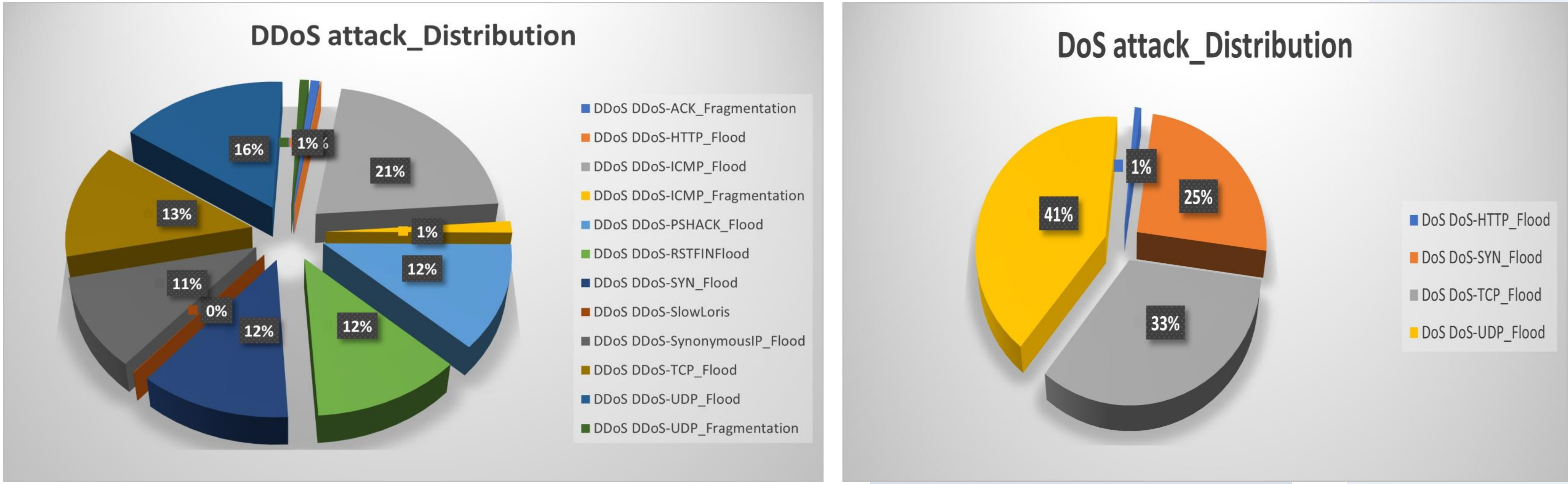
Percentage of the occurrence of each of the Mirai attacks
Data Utilized: Attack type and count



- ✓ **Primary Observation:** Mirai attack Greeth Flood ranked as no.1 most observed attacks, followed by Udpplain and Greip Flood

Chart 6: Comparison of attack distributions for DDoS vs. Dos

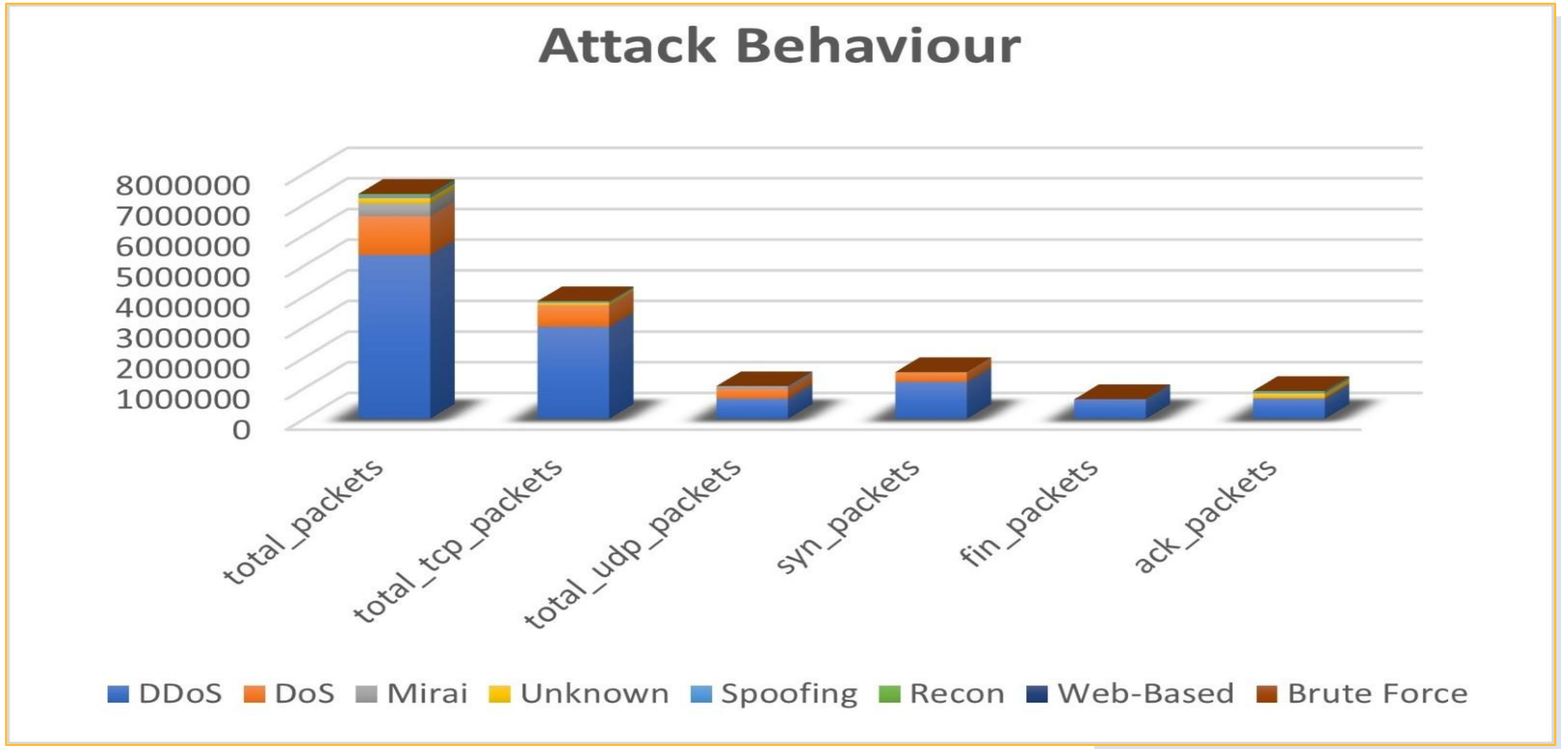
Percentage Distribution of Targeted Attacks by Category and Type
Data Utilized: Attack Category, label & total count



- ✓ **Primary Observation:** Among these DDoS attack_Distribution & DoS attack_Distrubution the DoS-UDP Flood was the most frequent in DoS attacks while DDoS_UDP_Flood was least attack in DDoS attacks.
- ✓ **Less Frequent Attacks:** DDoS-RSTFINFlood, DDoS_UDP_Fragmentation,DDoS-ICMP_Fragmentation are 1% attacks in DDoS attack, while only DoS-HTTP_Flood is the only 1% attack in DoS attack_Distribution.

Chart 7: Attack Behavior

Data Utilized: Attack category, total packets, total_tcp_packets, total_udp_packets, syn_packets, fin_packets and ack_packets

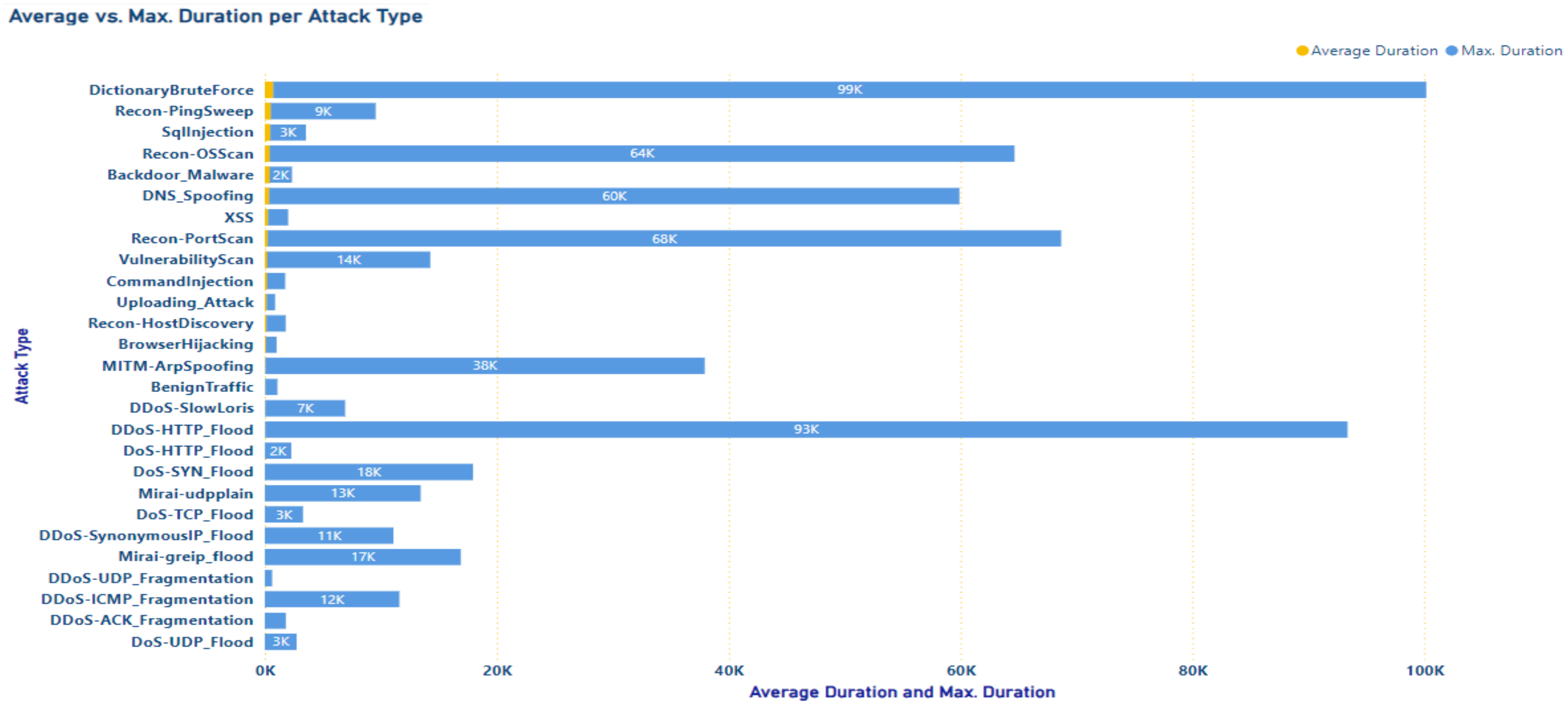


- ✓ **Primary Observation:** higher volume of total packets and total TCP packets associated with DDoS and DoS attacks. This indicates that these attack types are characterized by high levels of network traffic.
- ✓ **Less Targeted Networks:** Web based & brute force is the less target attack category
- ✓ **SYN, FIN, and ACK packets** are relatively less frequent across all attacks except for in the context of TCP-based attacks like DDoS and DoS, indicating that these packet flags are primarily significant in TCP traffic.

Chart 8: Duration Analysis per Attack Type

Average vs. Max. Duration per Attack

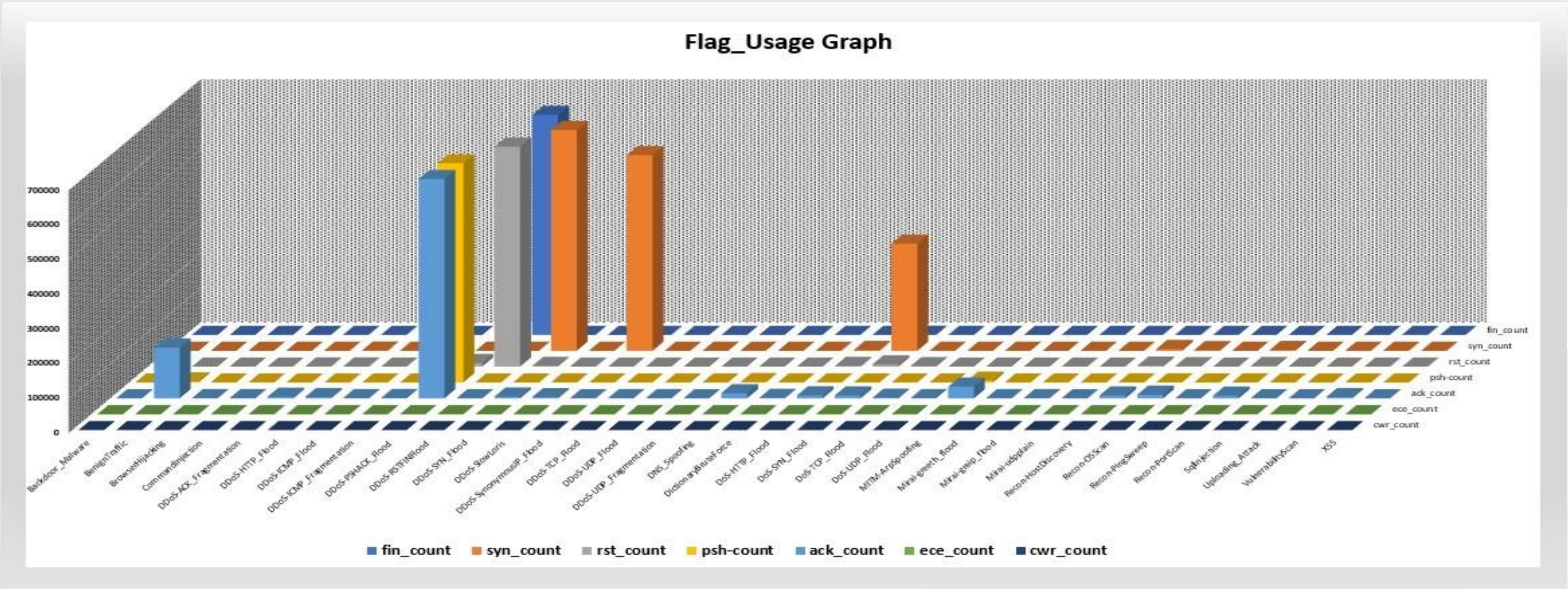
Data Utilized: Attack type, average and maximum duration



- ✓ **Primary Observation:** Certain types of attacks, such as "DictionaryBruteForce" and "MITM-ArpSpoofing," show notably long maximum durations, indicating that these attacks can persist for extended periods.

Chart 9: Flag Analysis (Packet Flow) per Attack

Data Utilized: Attack type, flags (fin, syn, psh, ack and ece), fin_count, syn_count, psh_count, ack_count and ece_count



Primary Observation:

- ✓ Attacks like "DDoS-TCP-Flood" and "BenignTraffic" show a very high usage of FIN and SYN flags, indicating that these attacks or conditions involve initiating or terminating a large number of TCP connections.
- ✓ Certain attacks, such as "DDoS-TCP-Flood" and "DDoS-UDP-Flood", show a high frequency of specific flags, particularly SYN and ACK flags. This suggests that these flags are heavily utilized in flooding attacks

Challenges and Solutions

Difficulties in handling data size after uploading the data in Hive .

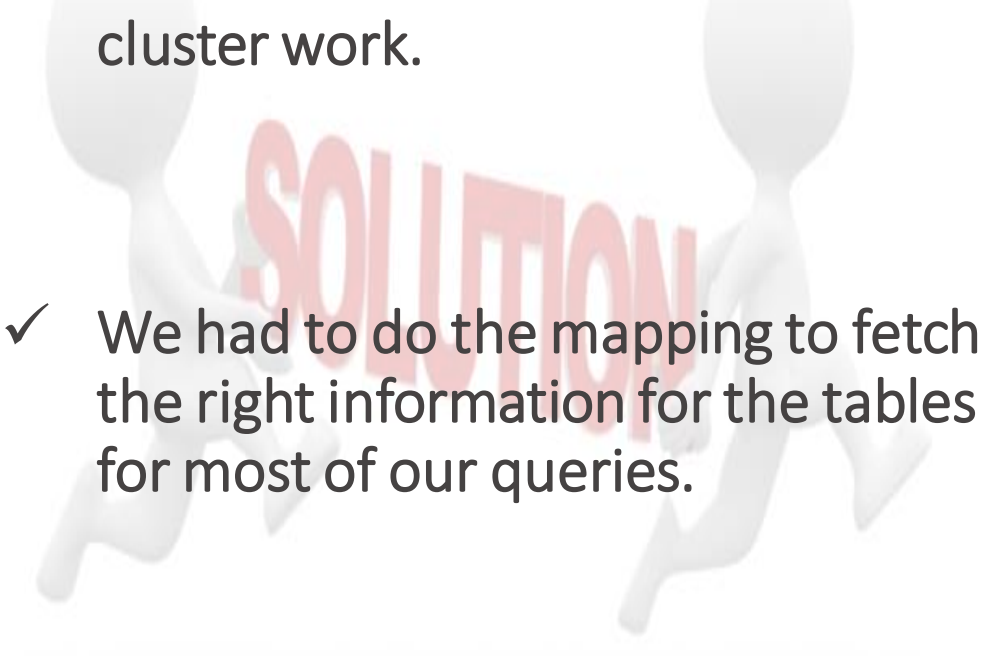
✓ The dataset was huge i.e., 12.5GB. We reduced it to 2GB to let the cluster work.

Data for Protocols & flags columns were in 0 & 1.

✓ We had to do the mapping to fetch the right information for the tables for most of our queries.

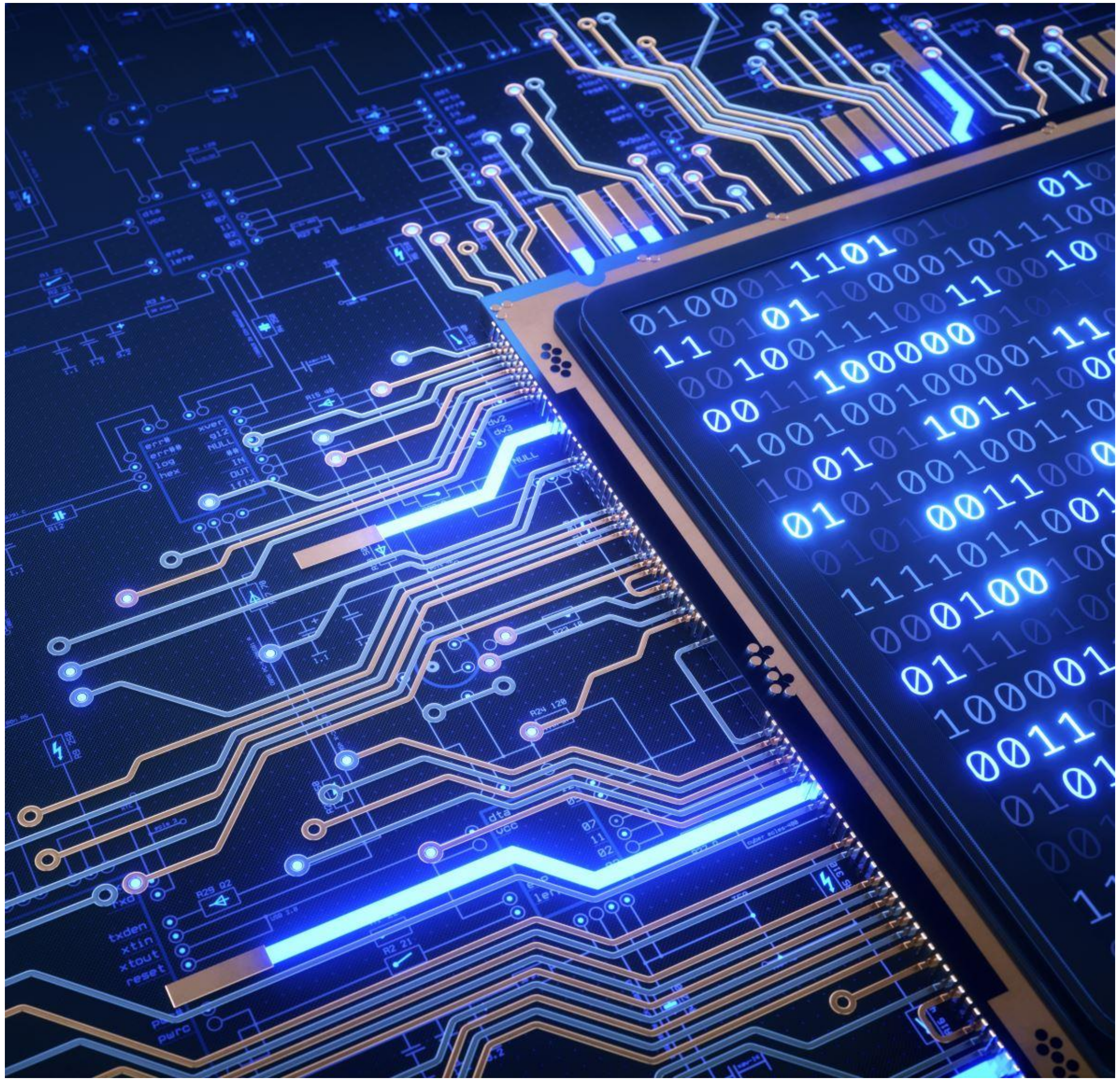
Table data was more of a numeric value. Hence hard to understand.

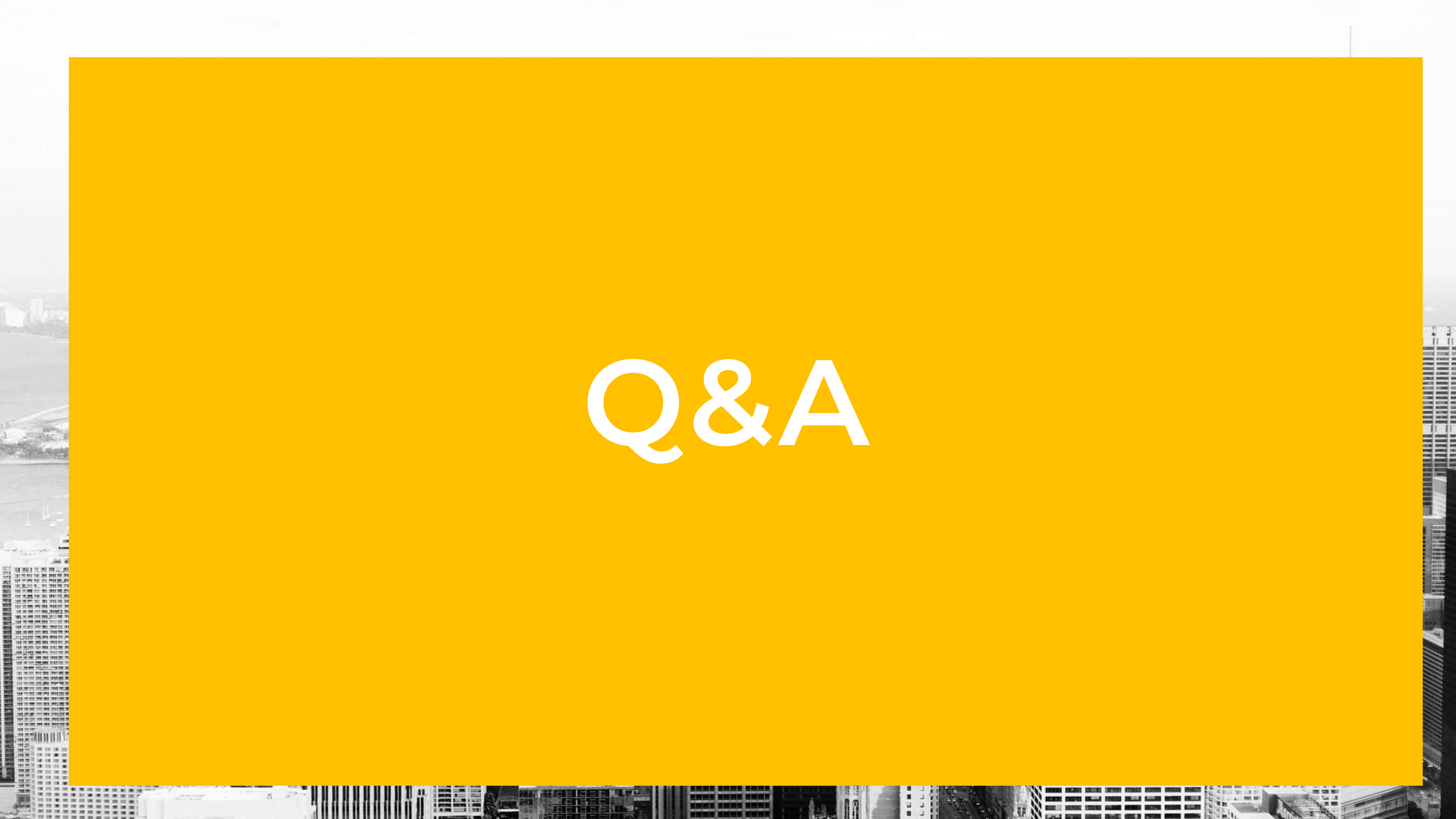
✓ Identified the columns which can build relations for all the queries in the analysis.



Conclusion

- The project effectively uses a rich dataset to provide a granular view of network attack dynamics in an IoT context. The findings not only enhance current understanding but also contribute to the development of more effective detection tools and defensive measures, ensuring that IoT environments are better protected against evolving cybersecurity threats. This strategic insight is crucial for anyone involved in IoT device security, network administration, or cybersecurity policy-making.





Q&A

THANK YOU