Network Traffic Analysis for Intrusion Detection

Presentation in Data Science for CyberSecurity and Forensics

DSCSF TY Btech. CSE (CSF) Case Study

Our team

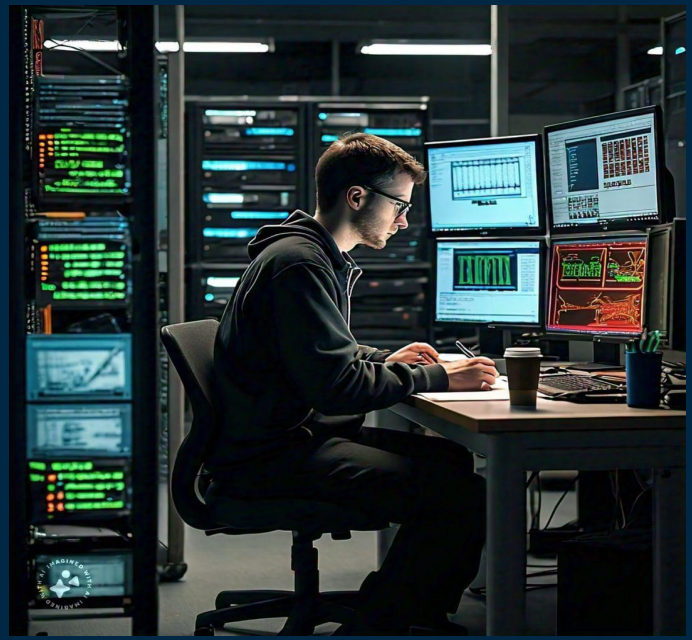
PA02 Mayur Behere	1032210568
PA10 Krishnaraj Thadesar	1032210888
PA23 Prathamesh Patil	1032211142
PA24 Saubhagya Singh	1032211144
PA25 Sourab Karad	1032211150

PA02

Mayur Behere

Introduction to Network Traffic Analysis

Network traffic analysis refers to the process of monitoring, capturing, and analyzing data packets transmitted over a network. It involves examining the characteristics, patterns, and content of network traffic to gain insights into the behavior of network users, applications, and devices.

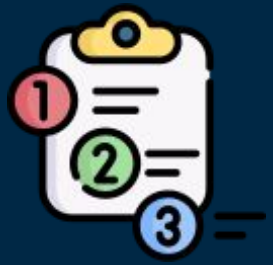


Importance of monitoring network traffic

Security

Intrusion Detection

Performance Optimization



Compliance Requirements

Troubleshooting

User Behavior Analysis

Goals of network traffic analysis for intrusion detection

1] Detection of Anomalies

2] Identification of Malicious Activities

3] Real-time Alerting

4] Incident Response Support

5] Threat Intelligence Gathering

6] Behavioral Profiling

PA23

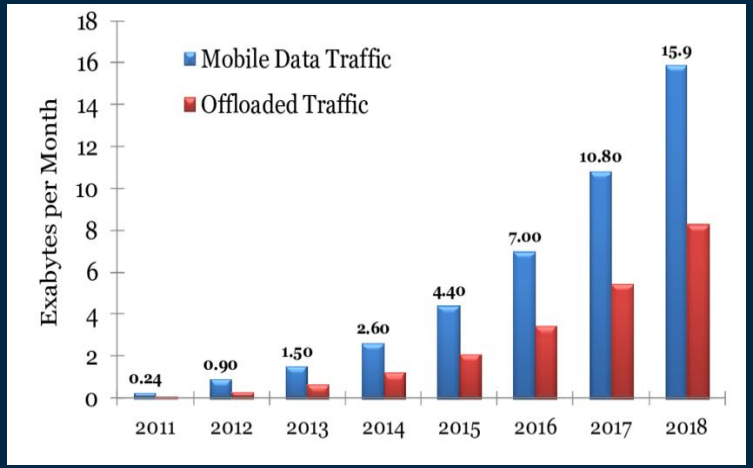
Prathamesh Patil

Types of Network Traffic

Data Traffic:

- TCP Traffic:
  Transmission Control Protocol (TCP) is used for reliable and connection-oriented communication. TCP traffic includes data transfers for applications like web browsing, email, file downloads/uploads, and database access.

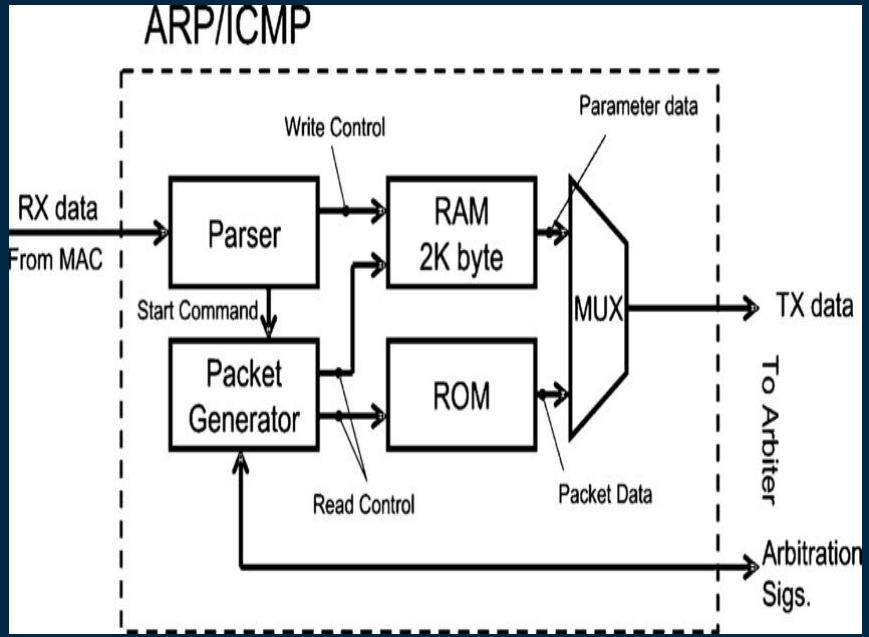
- UDP Traffic:
  User Datagram Protocol (UDP) is used for connectionless and lightweight communication. UDP traffic includes real-time applications such as VoIP (Voice over Internet Protocol), video streaming, online gaming, and DNS (Domain Name System) queries.



Control Traffic:

ICMP Traffic:
Internet Control Message Protocol (ICMP) is used for network control and error messaging. ICMP traffic includes ping requests, traceroute, and network error notifications.

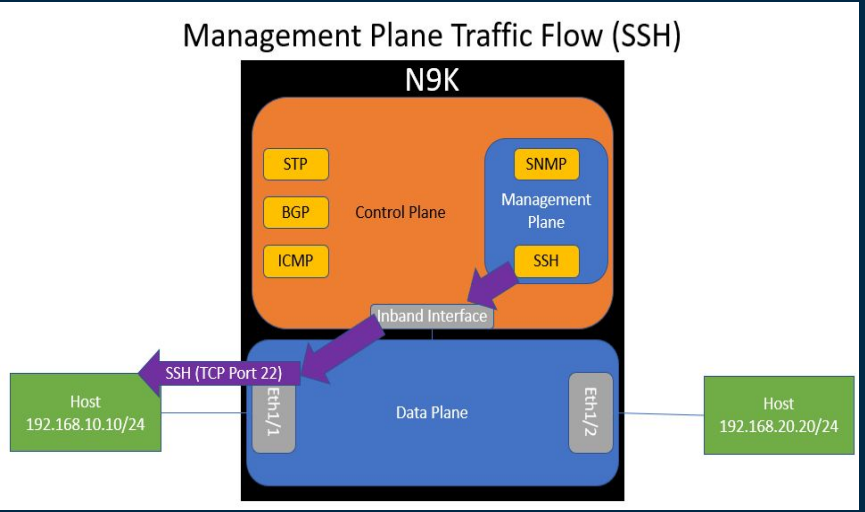
ARP Traffic:
Address Resolution Protocol (ARP) is used for mapping IP addresses to MAC addresses in local networks. ARP traffic includes ARP requests and responses for device communication within the same subnet.



Management Traffic:

SNMP Traffic: Simple Network Management Protocol (SNMP) is used for network monitoring and management. SNMP traffic includes queries and responses for monitoring network devices, collecting performance data, and managing configurations.

SSH Traffic: Secure Shell (SSH) traffic is used for secure remote access and management of network devices. SSH traffic includes encrypted login sessions, file transfers, and command executions.



Application-specific Traffic

HTTP Traffic
Hypertext Transfer Protocol

FTP Traffic
File Transfer Protocol

DNS Traffic
Domain Name System

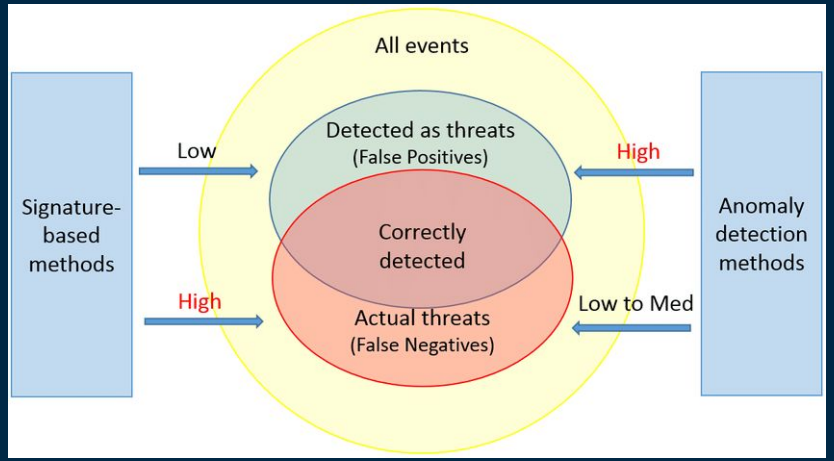
SMTP Traffic
Simple Mail Transfer Protocol

PA25
Sourab Karad

Traffic Analysis Techniques

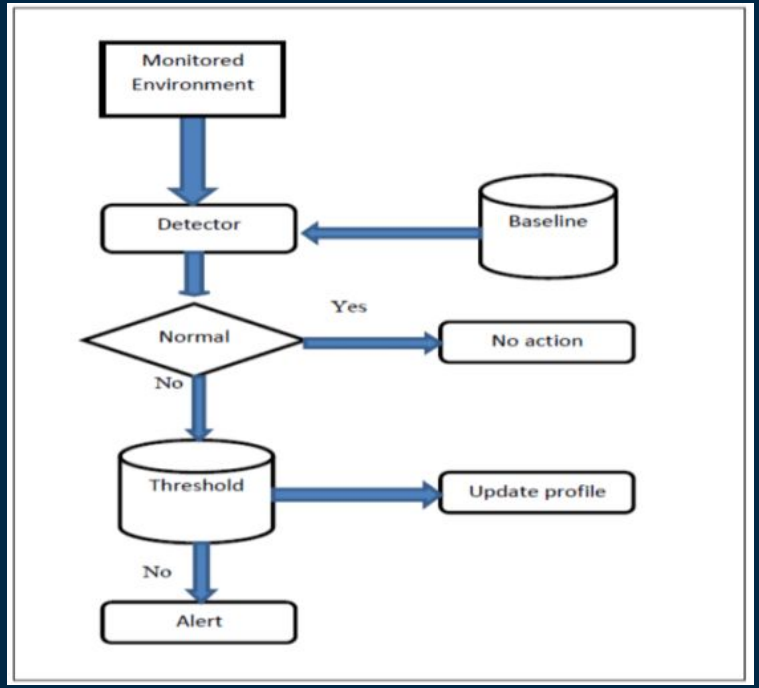
1. **Signature-Based Analysis:**

- Definition: Signature-based analysis involves comparing network traffic patterns against known attack signatures or predefined rules.
- Usage: It is effective for detecting well-known attacks with distinct signatures, such as known malware patterns, command and control communications, and specific exploit attempts.
- Tools: IDS/IPS systems with signature databases, Snort, Suricata.



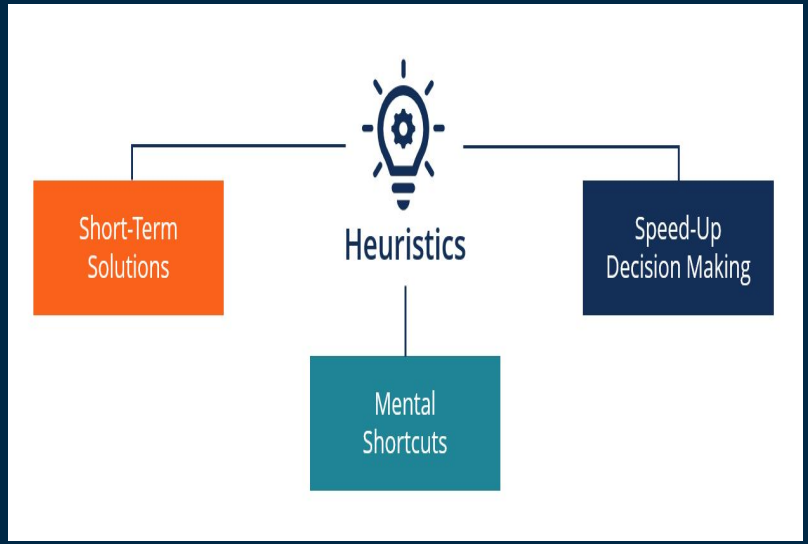
2. Anomaly-Based Analysis:

- Definition: Anomaly-based analysis focuses on identifying deviations from normal network behavior or traffic patterns.
- Usage: It is useful for detecting unknown or zero-day attacks, insider threats, and unusual network activities that do not match typical baseline behavior.
- Tools: Anomaly detection systems (ADS), machine learning models, NetFlow analyzers.



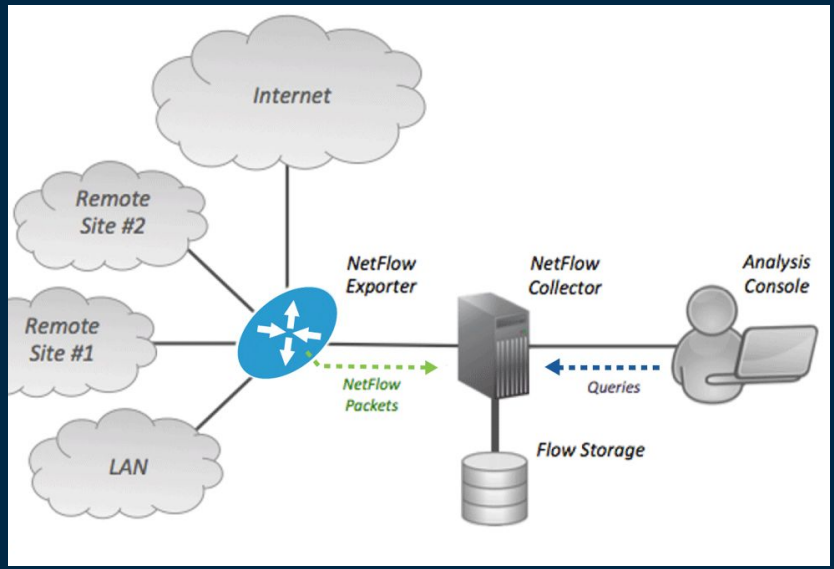
3. Heuristic-Based Analysis:

- Definition: Heuristic-based analysis relies on predefined heuristics or rules to detect suspicious or abnormal network behavior.
- Usage: It complements signature-based and anomaly-based techniques by targeting specific behaviors or characteristics indicative of potential threats.
- Tools: Custom scripts, network security appliances with heuristic capabilities.



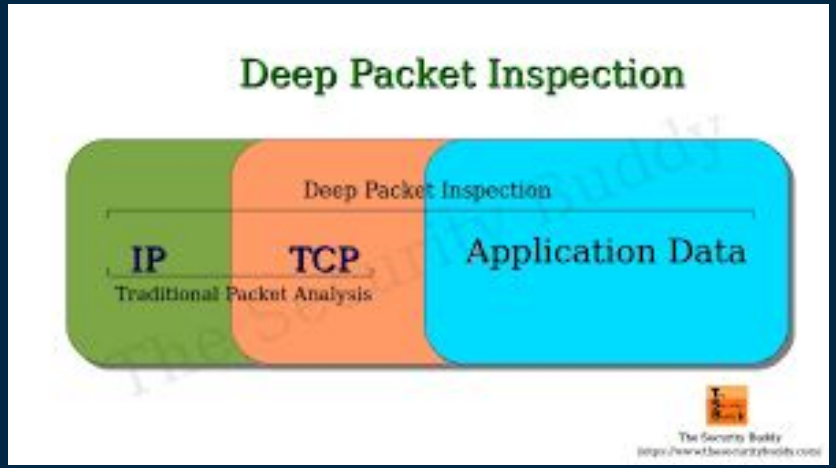
4. Flow-Based Analysis:

- Definition: Flow-based analysis focuses on analyzing network flow records, which provide summarized information about network communications between devices.
- Usage: It helps monitor traffic volumes, bandwidth utilization, communication patterns, and detect trends or anomalies based on flow data.
- Tools: NetFlow collectors, flow analyzers, SIEM platforms with flow analysis capabilities.



5. Deep Packet Inspection (DPI):

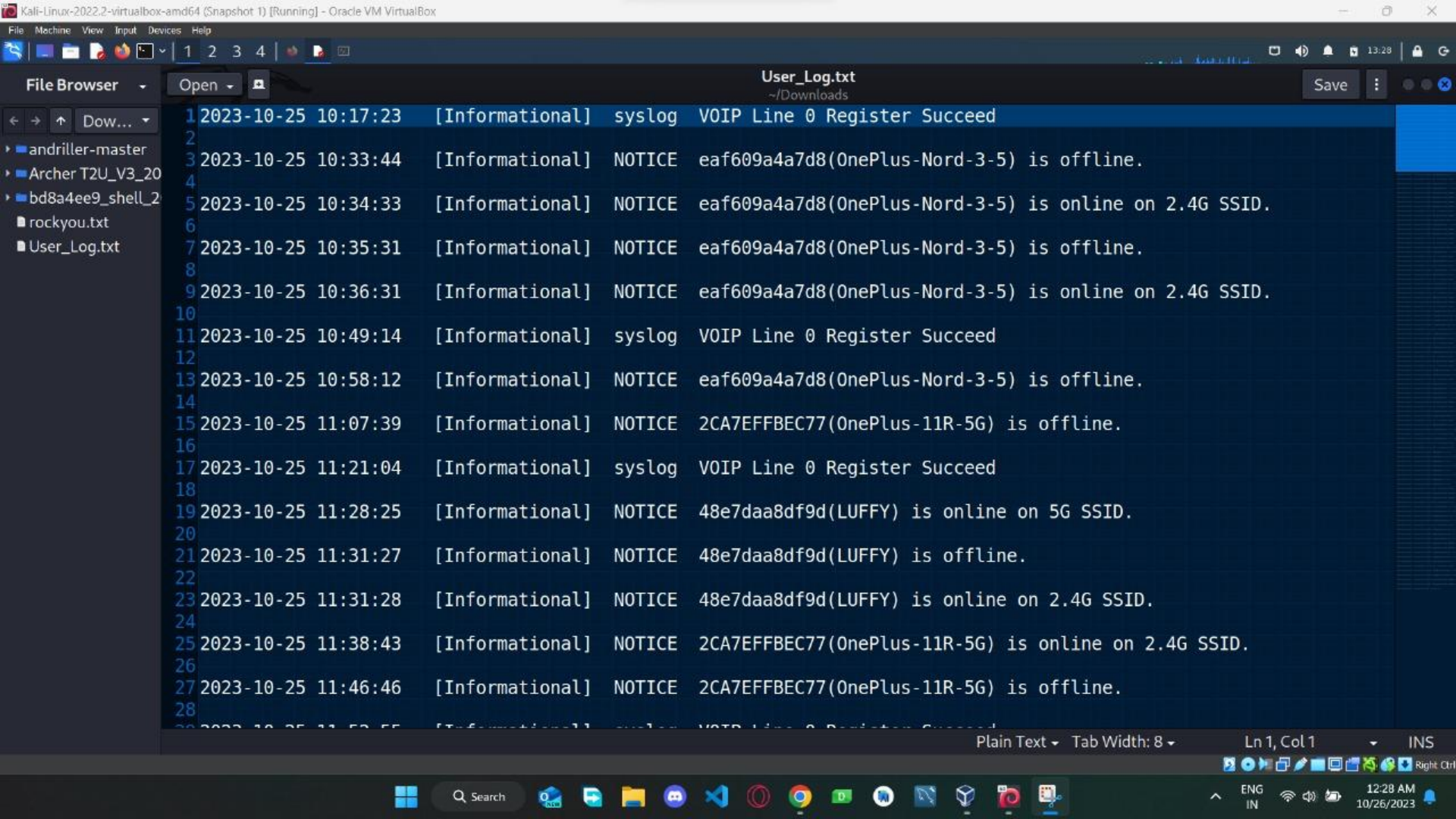
- Definition: DPI involves inspecting the contents of network packets at a granular level, including payload data, headers, and protocol information.
- Usage: It enables detailed analysis of application-layer protocols, content filtering, malware detection, and identification of malicious payloads.
- Tools: DPI-enabled firewalls, IDS/IPS systems, network traffic analyzers.

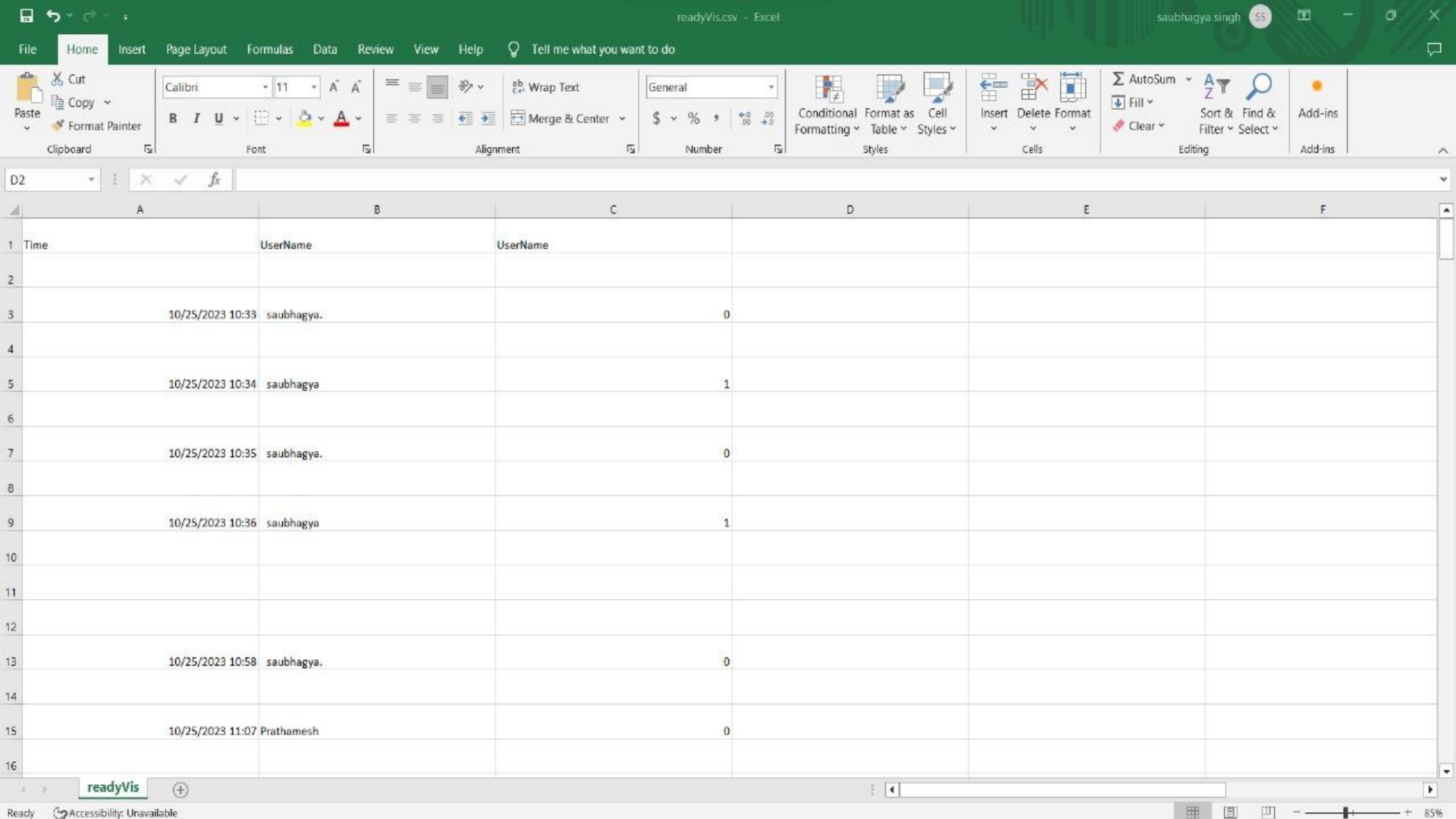


PA24

Saubhagya Singh

Data Collection and Processing


Kali-Linux-2022.2-virtualbox-amd64 (Snapshot 1) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
File Browser
Open
User_Log.txt
~/Downloads
Save
andriller-master
Archer T2U_V3_20
bd8a4ee9_shell_2
rockyou.txt
User_Log.txt
2023-10-25 10:17:23 [Informational] syslog VOIP Line 0 Register Succeed
2023-10-25 10:33:44 [Informational] NOTICE eaf609a4a7d8(OnePlus-Nord-3-5) is offline.
2023-10-25 10:34:33 [Informational] NOTICE eaf609a4a7d8(OnePlus-Nord-3-5) is online on 2.4G SSID.
2023-10-25 10:35:31 [Informational] NOTICE eaf609a4a7d8(OnePlus-Nord-3-5) is offline.
2023-10-25 10:36:31 [Informational] NOTICE eaf609a4a7d8(OnePlus-Nord-3-5) is online on 2.4G SSID.
2023-10-25 10:49:14 [Informational] syslog VOIP Line 0 Register Succeed
2023-10-25 10:58:12 [Informational] NOTICE eaf609a4a7d8(OnePlus-Nord-3-5) is offline.
2023-10-25 11:07:39 [Informational] NOTICE 2CA7EFFBEC77(OnePlus-11R-5G) is offline.
2023-10-25 11:21:04 [Informational] syslog VOIP Line 0 Register Succeed
2023-10-25 11:28:25 [Informational] NOTICE 48e7daa8df9d(LUFFY) is online on 5G SSID.
2023-10-25 11:31:27 [Informational] NOTICE 48e7daa8df9d(LUFFY) is offline.
2023-10-25 11:31:28 [Informational] NOTICE 48e7daa8df9d(LUFFY) is online on 2.4G SSID.
2023-10-25 11:38:43 [Informational] NOTICE 2CA7EFFBEC77(OnePlus-11R-5G) is online on 2.4G SSID.
2023-10-25 11:46:46 [Informational] NOTICE 2CA7EFFBEC77(OnePlus-11R-5G) is offline.
Plain Text
Tab Width: 8
Ln 1, Col 1
INS
Search
ENG
IN
12:28 AM
10/26/2023


readyVis.csv - Excel
saubhagya singh
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Help
Tell me what you want to do
Cut
Copy
Format Painter
Paste
Clipboard
Calibri
11
Font
Wrap Text
Merge & Center
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Editing
Add-ins
Add-ins
D2
A
B
C
D
E
F
1 Time UserName UserName
3 10/25/2023 10:33 saubhagya. 0
5 10/25/2023 10:34 saubhagya 1
7 10/25/2023 10:35 saubhagya. 0
9 10/25/2023 10:36 saubhagya 1
13 10/25/2023 10:58 saubhagya. 0
15 10/25/2023 11:07 Prathamesh 0
readyVis
Ready
Accessibility: Unavailable
85%

In [9]:
```
import matplotlib.pyplot as plt

# Create a timeline plot
plt.figure(figsize=(12, 6))

# Define a mapping of user names to y-coordinate positions
user_positions = {user: index for index, user in enumerate(df1_fifty['UserName'].unique())}

# Plot each user's activity
for user, user_data in df1_fifty.groupby('UserName'):
    user_data = user_data[user_data['UserName.1'] == 1]  # Filter for activity
    for _, row in user_data.iterrows():
        plt.plot([row['Time'], row['Time']], [user_positions[user], user_positions[user] + 1], color='blue')

# Set y-ticks and labels based on user positions
plt.yticks(list(user_positions.values()), user_positions.keys())  # Use yticks, not yticklabels

plt.xlabel('Time')
plt.title('Device Activity Timeline')
plt.grid(axis='x')
plt.tight_layout()

plt.show()
```

In [7]:
```
import matplotlib.pyplot as plt
df1_fifty=df1.head(30)
df1_fifty['Time'] = pd.to_datetime(df1_fifty['Time'])
```

```
C:\Users\saubhagya singh\AppData\Local\Temp\ipykernel_14100\1941238947.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df1_fifty['Time'] = pd.to_datetime(df1_fifty['Time'])
```

In [8]:
```
plt.figure(figsize=(20, 6))
plt.scatter(df1_fifty['UserName'],df1_fifty['UserName.1'], label='Data Points', color='blue', marker='o')

plt.title('Wifi Activity')
plt.xlabel('X-axis Label')
plt.ylabel('Y-axis Label')
plt.grid(True)
plt.show()
```

In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv('userlogs.csv')
```

In [3]:
```
df1=pd.read_csv('readyVis.csv')
```

In [4]:
```
df1
```

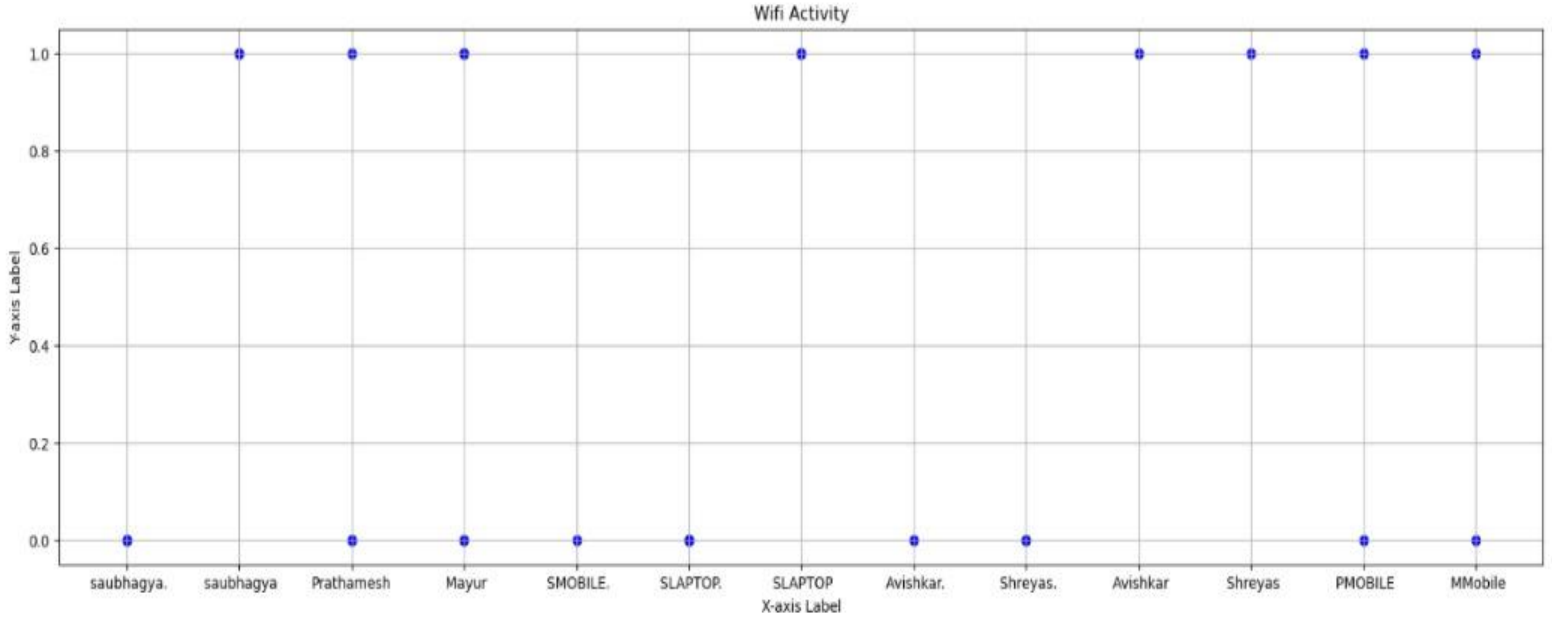
Out[4]:

	Time	UserName	UserName.1
0	NaN	NaN	NaN
1	10/25/2023 10:33	saubhagya.	0.0
2	NaN	NaN	NaN
3	10/25/2023 10:34	saubhagya	1.0
4	NaN	NaN	NaN
...	...	...	...
207	10/26/2023 0:18	SLAPTOP.	0.0
208	NaN	NaN	NaN
209	10/26/2023 0:18	SLAPTOP	1.0
210	NaN	NaN	NaN
211	NaN	NaN	NaN

212 rows × 3 columns

```
plt.scatter(df1_fifty['UserName'],df1_fifty['UserName.1'], label='Data Points', color='blue', marker='o')

plt.title('Wifi Activity')
plt.xlabel('X-axis Label')
plt.ylabel('Y-axis Label')
plt.grid(True)
plt.show()
```



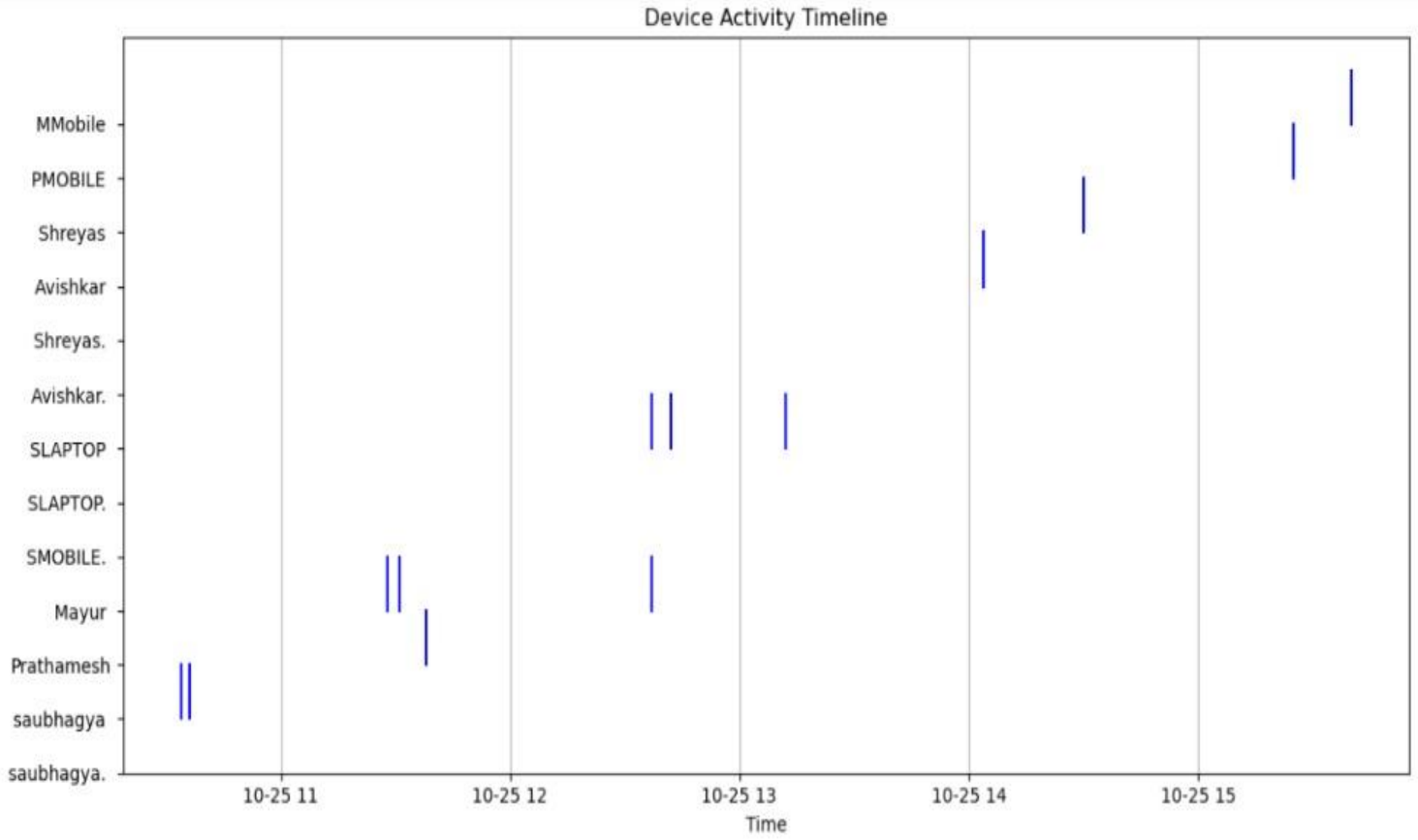
jupyter Wifi logs Last Checkpoint: a minute ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3 (ipykernel)

Run Code



Common Network Threats and Attack Patterns

1. Malware Attacks
2. Phishing and Social Engineering
3. Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS) Attacks
4. Man-in-the-Middle (MitM) Attacks
5. SQL Injection (SQLi) Attacks
6. Cross-Site Scripting (XSS) Attacks
7. Brute Force and Credential Stuffing Attacks
8. Insider Threats
9. IoT-Based Attacks
10. Data Exfiltration

Detection Strategies and Algorithms

1. Signature-Based Detection:
    - Description: Compares network traffic against known attack signatures or patterns.

2. Anomaly-Based Detection:
    - Description: Identifies deviations from normal network behavior.

3. Heuristic-Based Detection:
    - Description: Uses predefined rules or heuristics to detect suspicious activities.

4. Machine Learning (ML) Based Detection:
    - Description: Utilizes ML algorithms to learn and detect patterns indicative of threats.

PA10

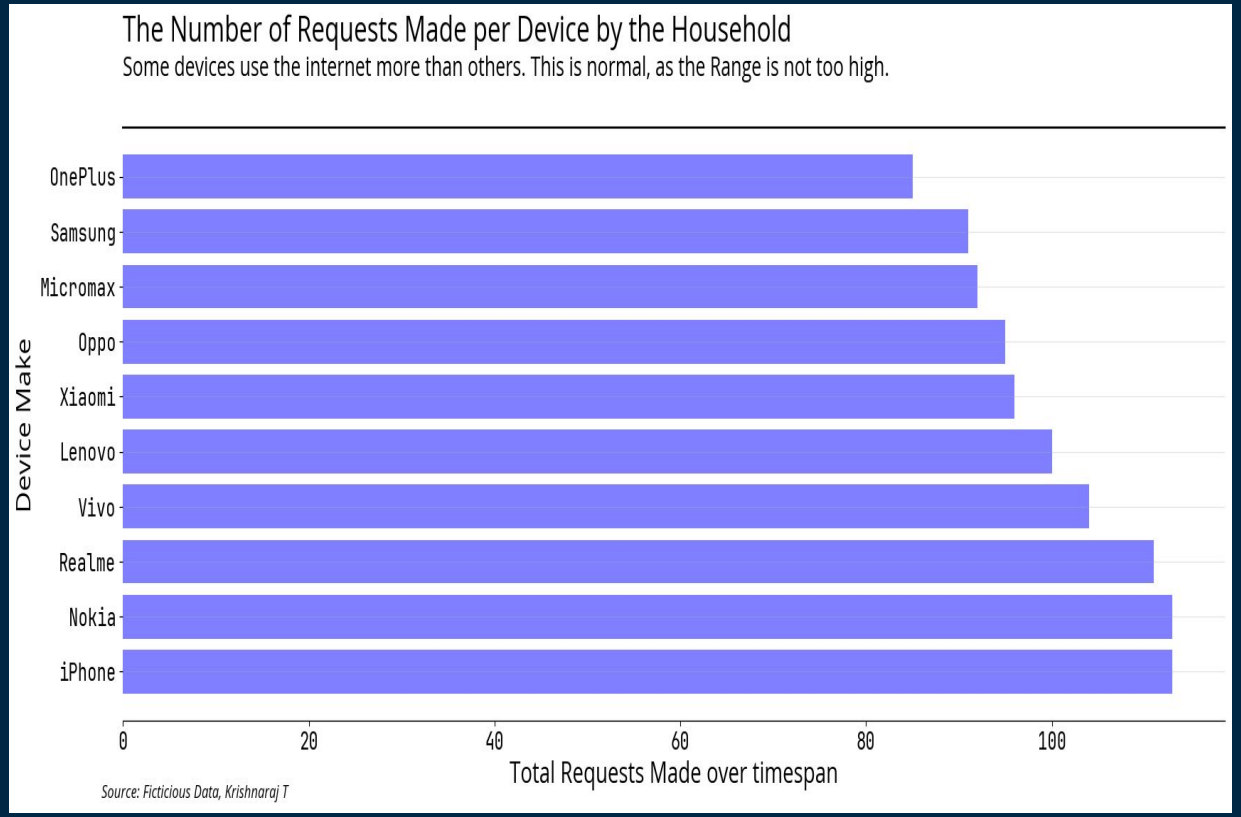
Krishnaraj Thadesar

Quatari

Case Studies and Examples

DOS Attack Data

From the graph, we can observe that OnePlus and Samsung devices seem to have made the highest number of requests, followed by Micromax, Oppo, and Xiaomi. On the other hand, iPhones appear to have made the fewest requests compared to the other device makes listed.



Out[5]:

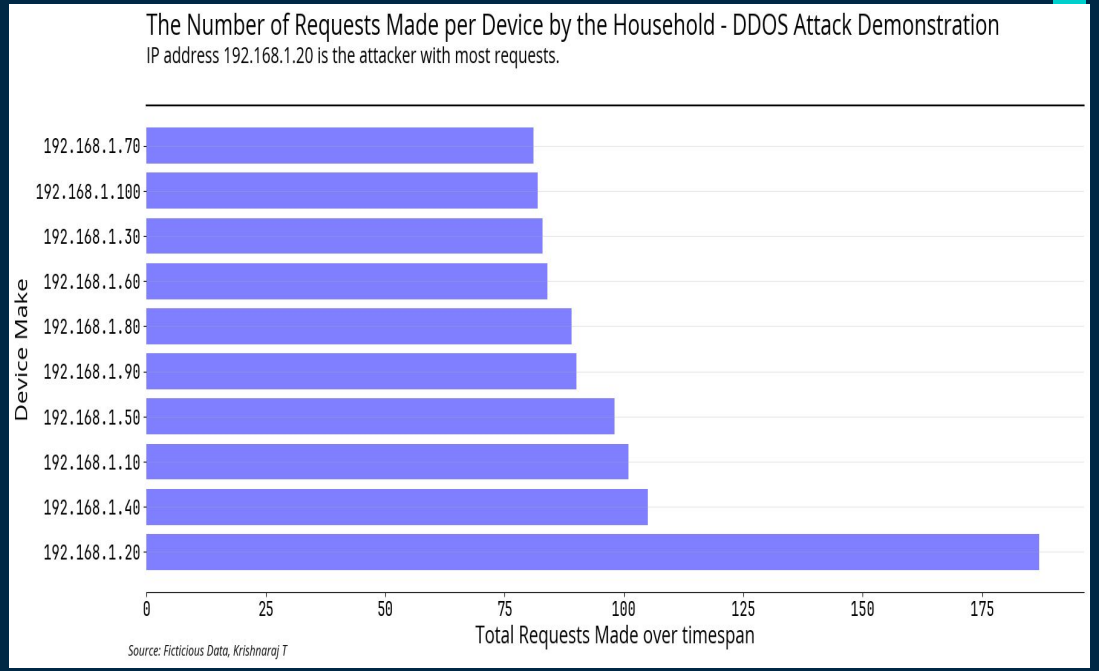
	MAC	IP Address	Device Name	Interface	Requested IP	Time	Requested Website	Protocol	Port
0	87:f1:20:92:23:58	192.168.1.60	iPhone	2.4gz	216.58.194.45	2023-01-09 08:39:54	Youtube	TCP	21.0
1	85:f7:d5:30:88:31	192.168.1.80	Oppo	2.4gz	69.63.176.22	2023-01-02 10:56:14	Facebook	HTTPS	21.0
2	bc:34:18:65:0a:ab	192.168.1.100	iPhone	2.4gz	3.213.31.34	2023-01-04 12:40:38	Facebook	HTTPS	53.0
3	7c:98:6a:ce:01:3d	192.168.1.90	iPhone	5gz	3.213.31.34	2023-01-05 10:11:18	Instagram	HTTPS	21.0
4	19:56:ab:0d:af:3f	192.168.1.100	Realme	2.4gz	104.244.42.12	2023-01-08 22:37:53	Youtube	TCP	21.0
...	...	...	...	...	...	...	...	...	...
995	1d:b8:e4:8c:cf:6f	192.168.1.20	Realme	2.4gz	3.213.31.34	2023-01-07 13:08:10	Other	TCP	67.0
996	06:b3:c5:e5:ca:3e	192.168.1.80	Micromax	2.4gz	3.213.31.34	2023-01-03 10:42:07	Other	HTTP	110.0
997	46:b9:89:c6:bc:0a	192.168.1.100	OnePlus	2.4gz	3.213.31.34	2023-01-01 16:18:56	Facebook	HTTP	53.0
998	d0:aa:4f:d9:17:a1	192.168.1.90	Realme	5gz	216.58.194.45	2023-01-04 20:35:42	Youtube	UDP	NaN
999	61:8a:3f:44:f2:d4	192.168.1.80	Xiaomi	2.4gz	216.58.194.45	2023-01-07 17:49:28	Instagram	HTTP	53.0

IP Address Connected– DOS Attack

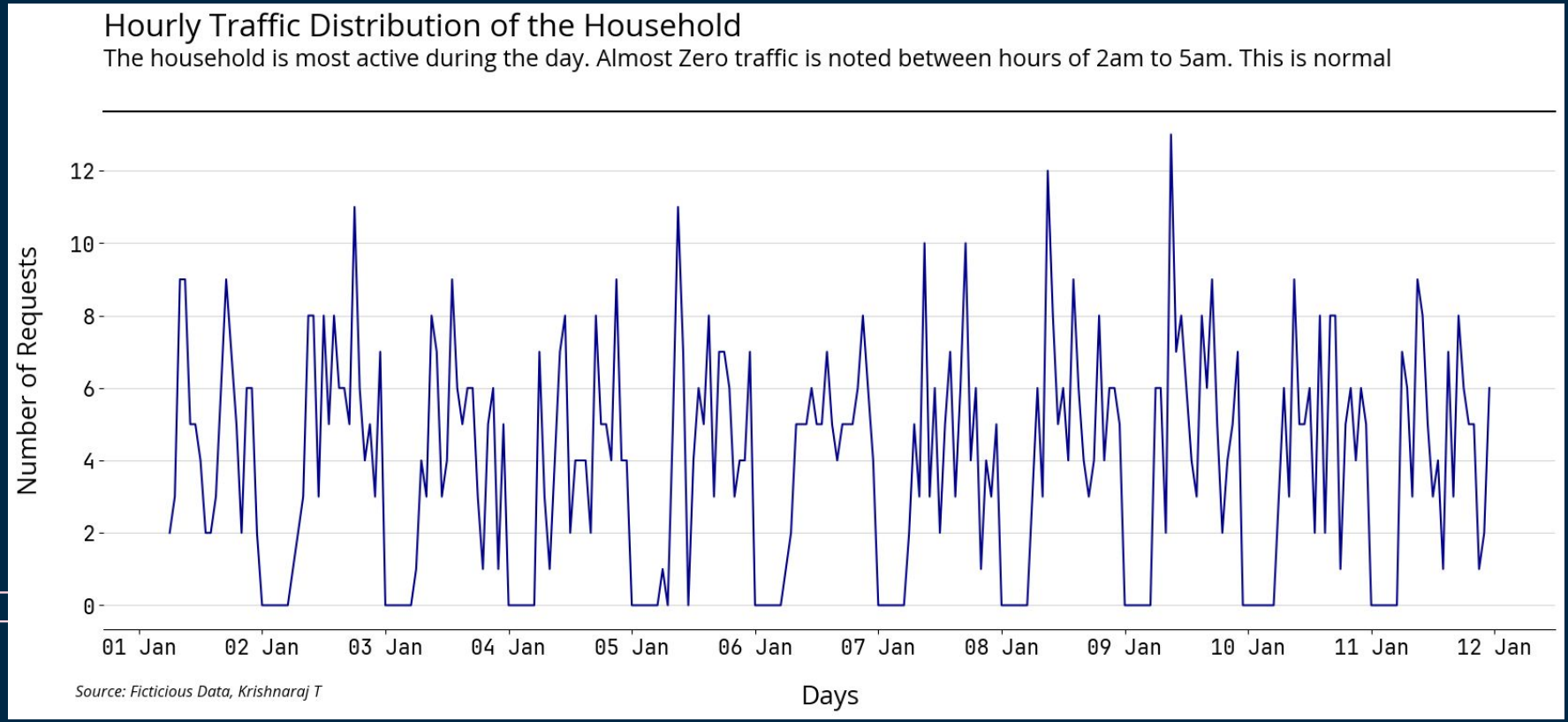
This bar graph illustrates "The Number of Requests Made per Device by the Household - DDOS Attack Demonstration." Each bar represents a different device identified by its IP address, and the length of the bar corresponds to the total number of requests made by that specific device over a certain time span.

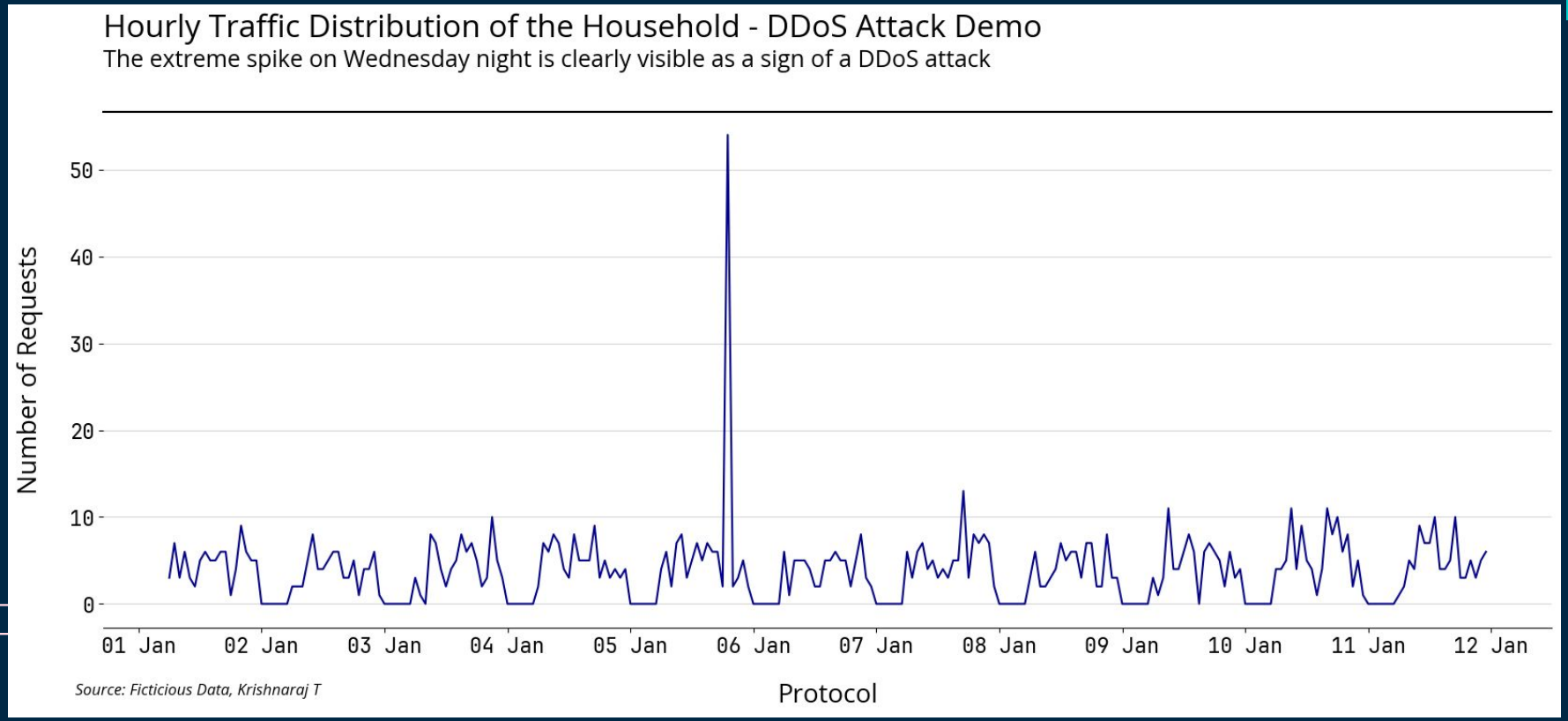
The graph highlights that the IP address 192.168.1.20 has made the most requests, indicating that this device is the attacker with the highest number of requests in a potential Distributed Denial of Service (DDOS) attack scenario.

Other IP addresses listed in the graph also show varying levels of requests made, with 192.168.1.70, 192.168.1.100, and 192.168.1.30 being among the top-requesting devices after the attacker.



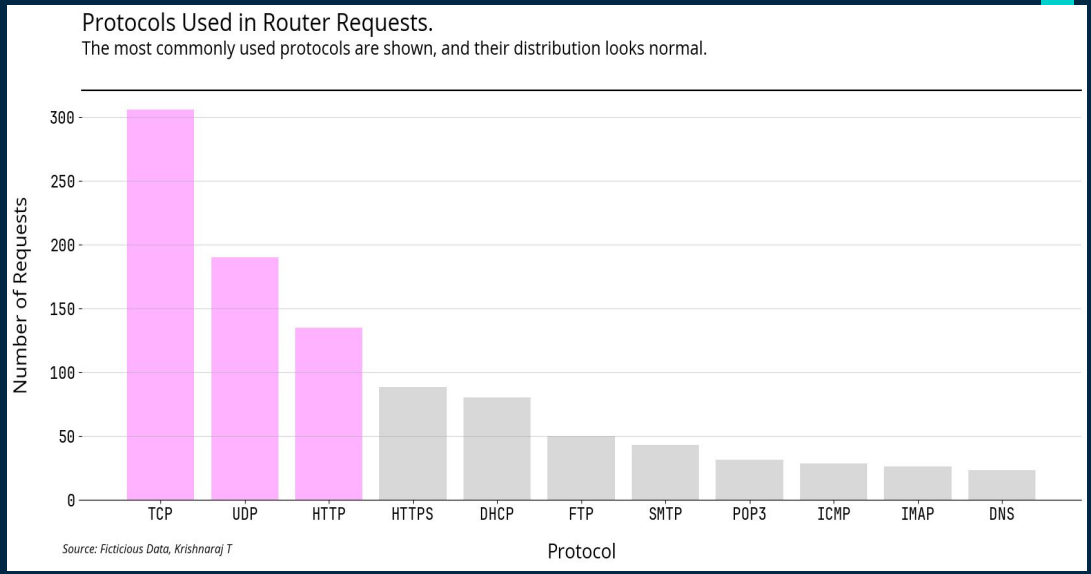
The fluctuating pattern of the graph indicates varying levels of internet usage throughout the day, with peaks and troughs representing periods of higher and lower activity, respectively. Overall, the graph provides insights into the hourly distribution of internet traffic within the household over the specified time period.



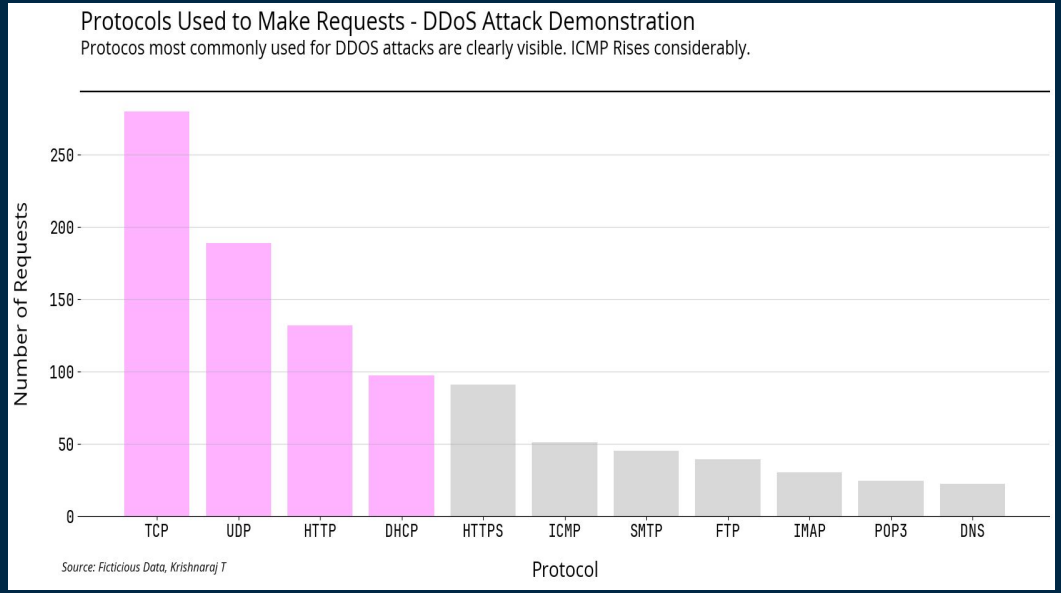

Hourly Traffic Distribution of the Household - DDoS Attack Demo
The extreme spike on Wednesday night is clearly visible as a sign of a DDoS attack
Number of Requests
50
40
30
20
10
0
01 Jan
02 Jan
03 Jan
04 Jan
05 Jan
06 Jan
07 Jan
08 Jan
09 Jan
10 Jan
11 Jan
12 Jan
Protocol
Source: Ficticious Data, Krishnaraj T

Router Request

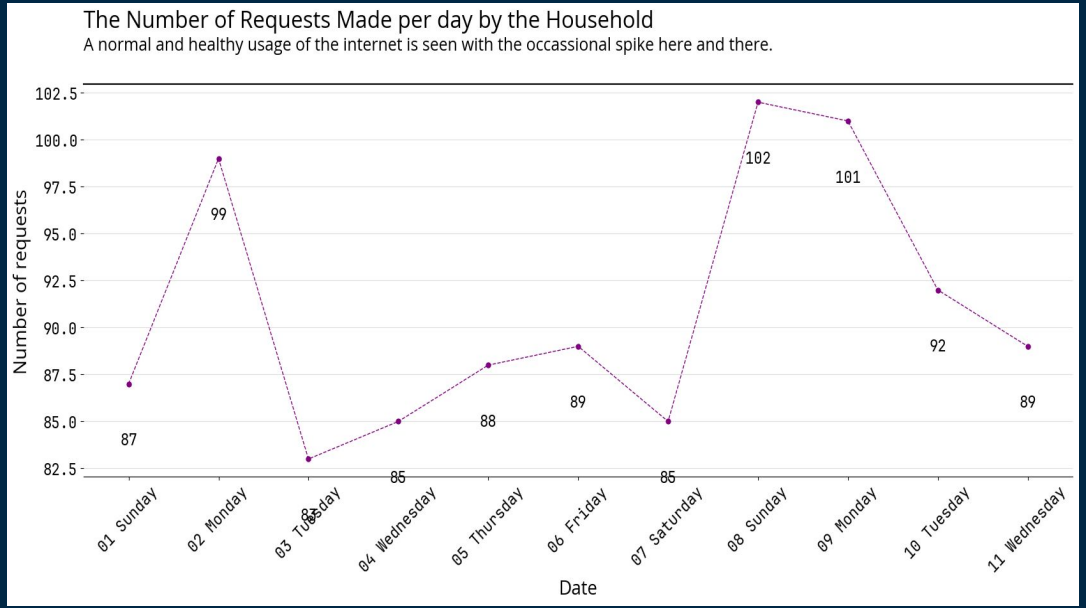
1. TCP (Transmission Control Protocol): It appears to be the most commonly used protocol, with the highest number of requests.
2. UDP (User Datagram Protocol): It's the second most used protocol, with a significant number of requests, although less than TCP.
3. HTTP (Hypertext Transfer Protocol): This protocol is also widely used, with a notable number of requests.
4. HTTPS (Hypertext Transfer Protocol Secure): HTTPS follows HTTP in terms of usage but with fewer requests.
5. DHCP (Dynamic Host Configuration Protocol): It shows a moderate level of usage compared to other protocols.
6. FTP (File Transfer Protocol), SMTP (Simple Mail Transfer Protocol), POP3 (Post Office Protocol version 3), ICMP (Internet Control Message Protocol), IMAP (Internet Message Access Protocol), DNS (Domain Name System): These protocols show lower levels of usage compared to the top protocols mentioned earlier.



This bar graph illustrates the "Protocols Used to Make Requests - DDOS Attack Demonstration." It highlights the protocols most commonly used for DDOS (Distributed Denial of Service) attacks, with a particular emphasis on the notable rise in ICMP (Internet Control Message Protocol) requests.

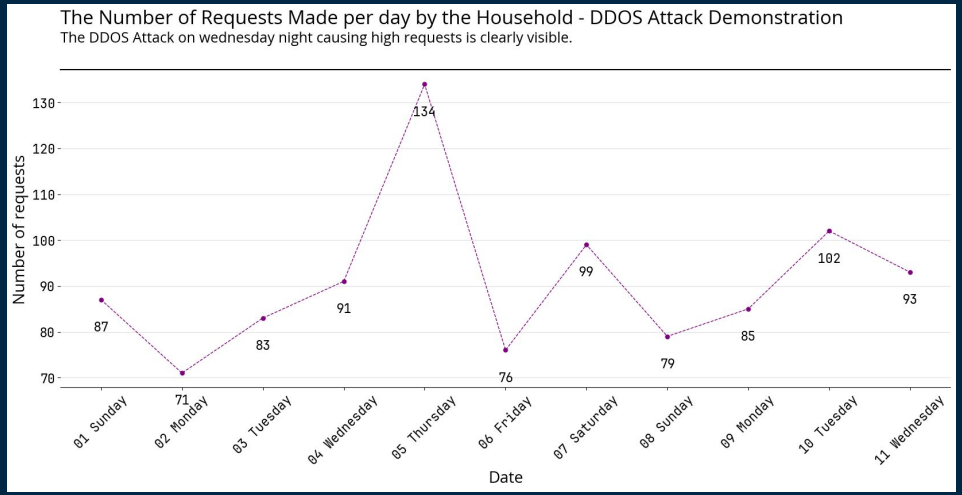


- X-Axis (Date): The dates from the 1st to the 11th of a month are displayed, indicating the timeline of the data.
- Y-Axis (Number of Requests): Represents the count of requests made by the household on each day. The values range from around 82.5 to 102.5.
- Data Points: Each data point on the graph represents the number of requests made on a specific day. The purple dots connected by dashed lines illustrate the trend in the number of requests over time.
- Annotations: Some data points are labeled with specific values, indicating the exact number of requests made on those particular days.



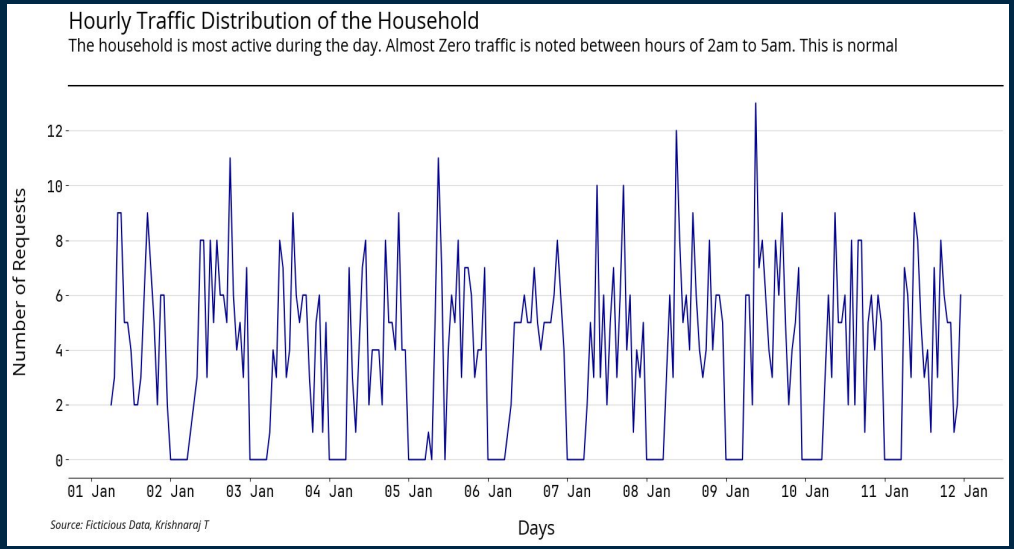
- "The DDOS Attack on Wednesday night causing high requests is clearly visible." offers additional information about the notable event depicted in the graph.

- Some data points are labeled with specific values, indicating the exact number of requests made on those particular days. Additionally, a text annotation highlights the occurrence of a DDoS attack on Wednesday night, which resulted in a significant increase in requests.



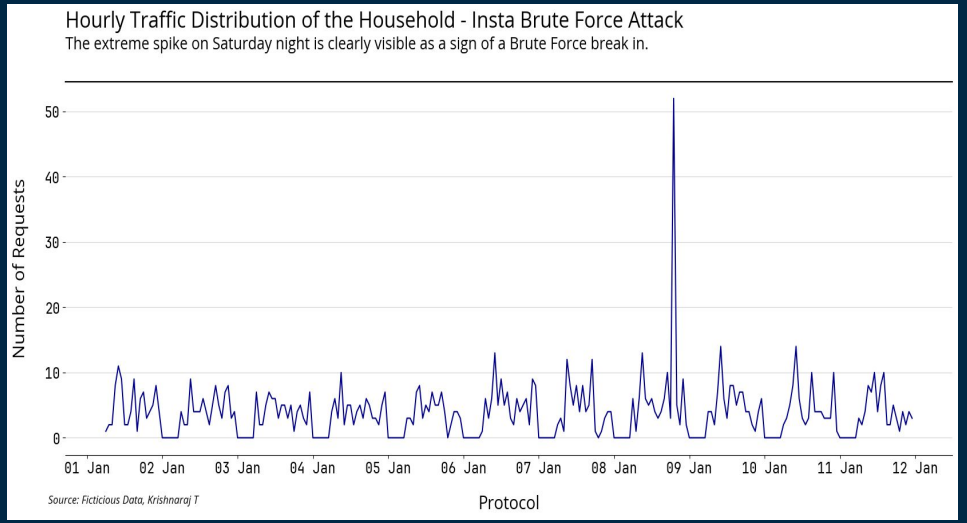
- "Hourly Traffic Distribution of the Household" provides the main subject of the graph, indicating that it depicts the traffic patterns of a household over time.

- "The household is most active during the day. Almost zero traffic is noted between hours of 2 am to 5 am. This is normal" offers additional information about the traffic patterns, highlighting that the household is most active during the day, with minimal activity during late-night hours, which is considered normal.



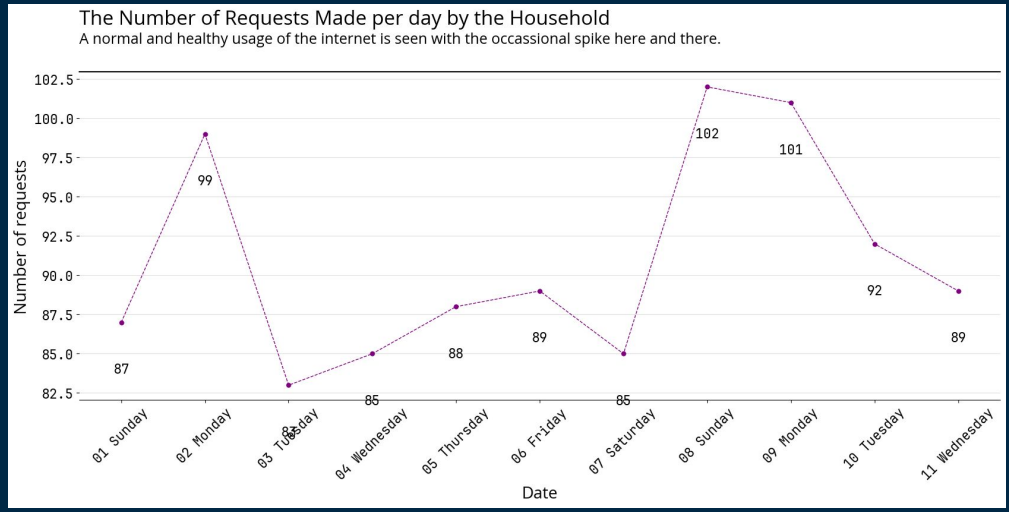
- "Hourly Traffic Distribution of the Household - Insta Brute Force Attack" suggests that the graph depicts the household's traffic patterns, specifically highlighting an incident of a brute force attack targeting Instagram.

- "The extreme spike on Saturday night is clearly visible as a sign of a Brute Force break-in." This provides additional context, indicating that the significant spike in traffic observed on Saturday night is indicative of a brute force attack on Instagram.



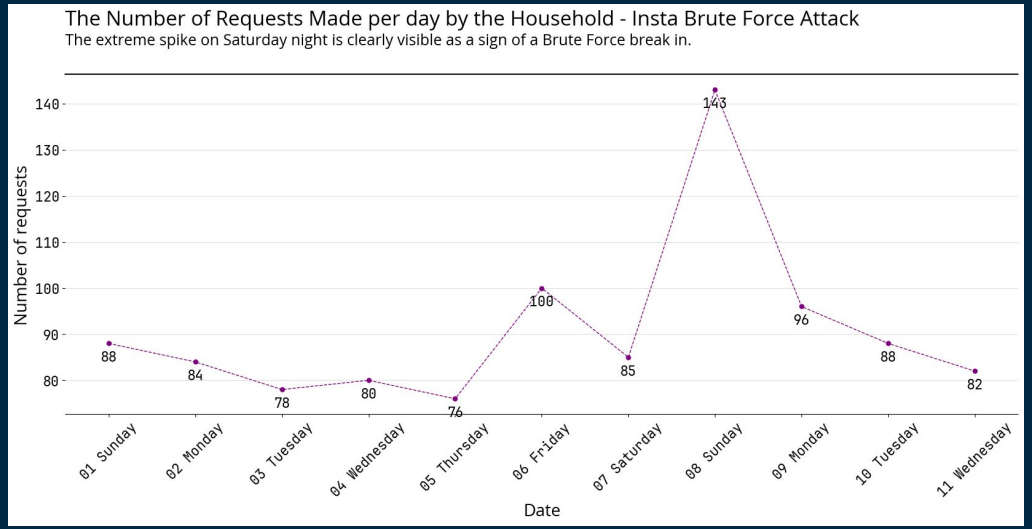
- "The Number of Requests Made per day by the Household" provides the main subject of the graph, indicating that it depicts the daily requests made by the household.

- "A normal and healthy usage of the internet is seen with the occasional spike here and there." This subtext suggests that the overall pattern of requests is within normal limits, with occasional spikes in activity.

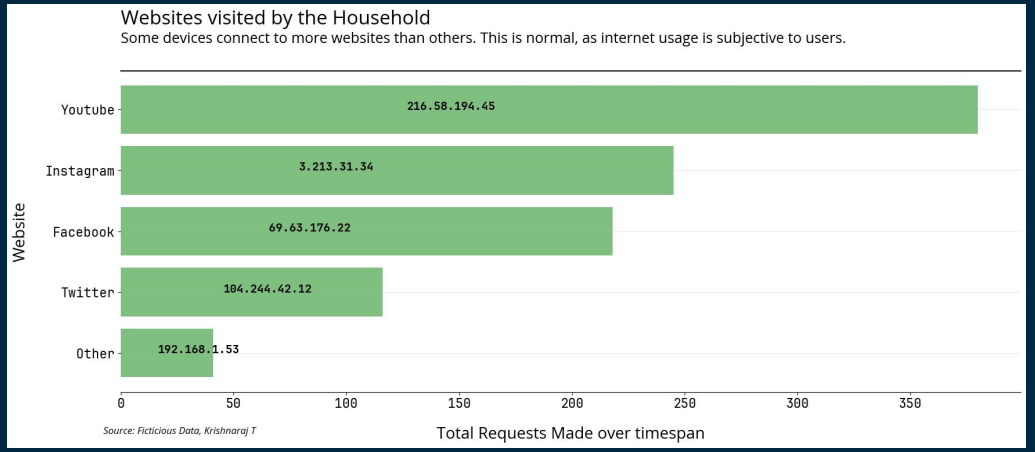


- "The Number of Requests Made per day by the Household - Insta Brute Force Attack" highlights the main subject of the graph, indicating that it depicts the daily requests made by the household, specifically during an event character.

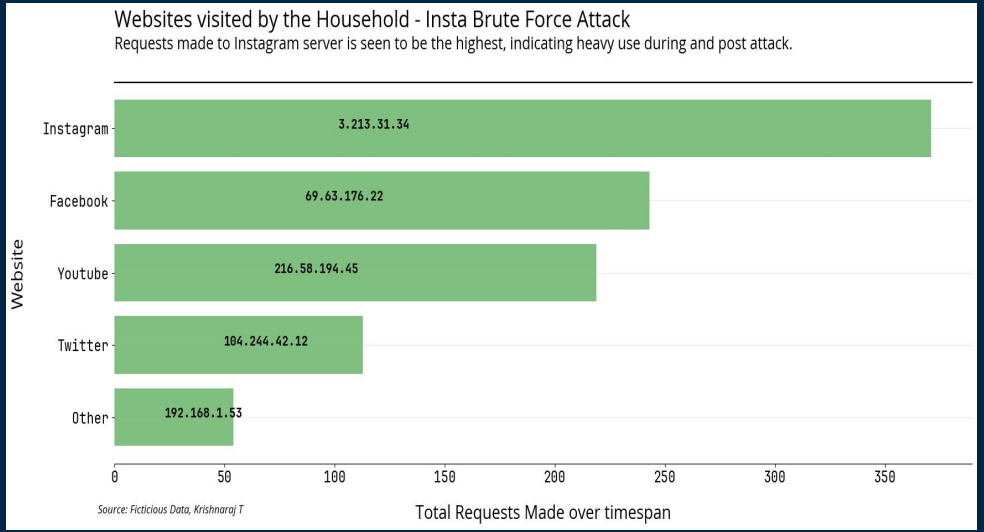
- "The extreme spike on Saturday night is clearly visible as a sign of a Brute Force break in." This subtext emphasizes that the drastic increase in requests observed on Saturday night is indicative of a Brute Force Attack.rized as a "Brute Force Attack."



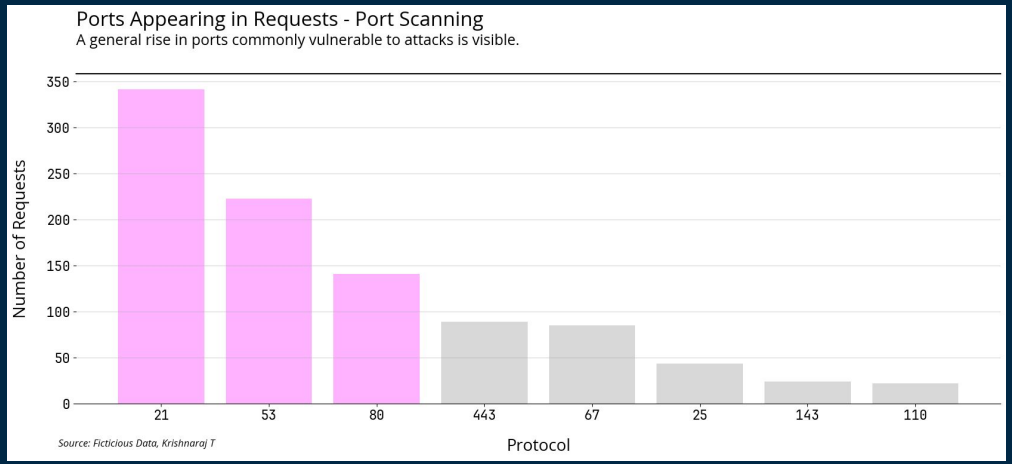
- "Websites visited by the Household" indicates the subject matter of the graph, which is the websites accessed by the household.
- "Some devices connect to more websites than others. This is normal, as internet usage is subjective to users." This statement suggests that different devices in the household may access varying numbers of websites, which is a normal aspect of internet usage



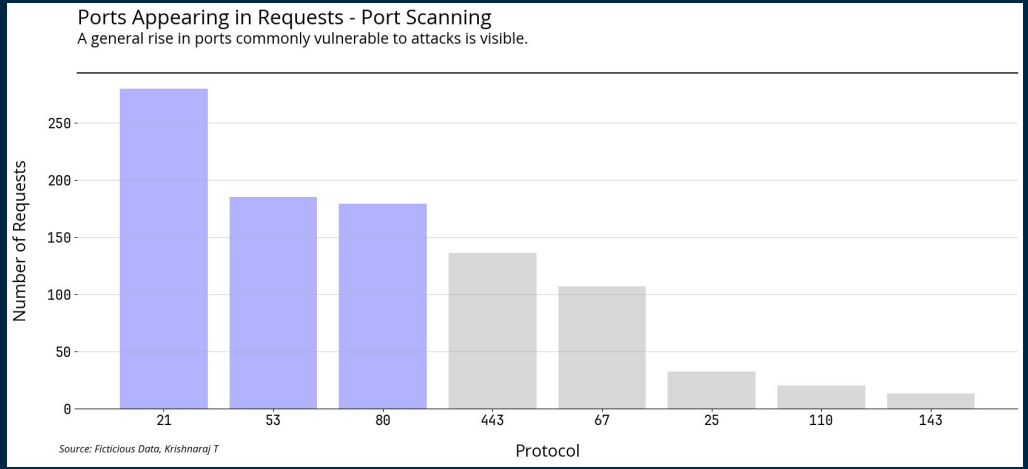
- "Websites visited by the Household - Insta Brute Force Attack" indicates that the data displayed pertains to website visits during an Insta Brute Force Attack.
- "Requests made to Instagram server is seen to be the highest, indicating heavy use during and post attack." This statement suggests that the Instagram server received the most requests during and after the Insta Brute Force Attack.



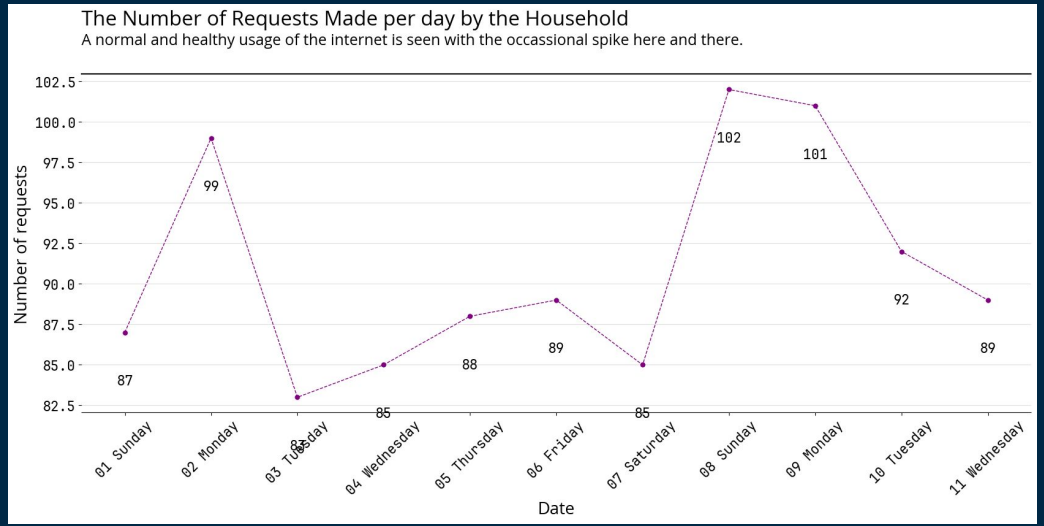
- "Ports Appearing in Requests - Port Scanning" indicates that the graph represents the appearance of different ports in requests, suggesting potential port scanning activity.
- "A general rise in ports commonly vulnerable to attacks is visible." This statement suggests that there's an increase in requests targeting ports that are commonly vulnerable to attacks.
- Ports 21 (FTP), 53 (DNS), 80 (HTTP), and 443 (HTTPS) have relatively higher bars, suggesting a higher number of requests made to access these ports.
- Ports 67 (DHCP), 25 (SMTP), 143 (IMAP), and 110 (POP3) have relatively lower bars, indicating a lower number of requests made to access these ports.



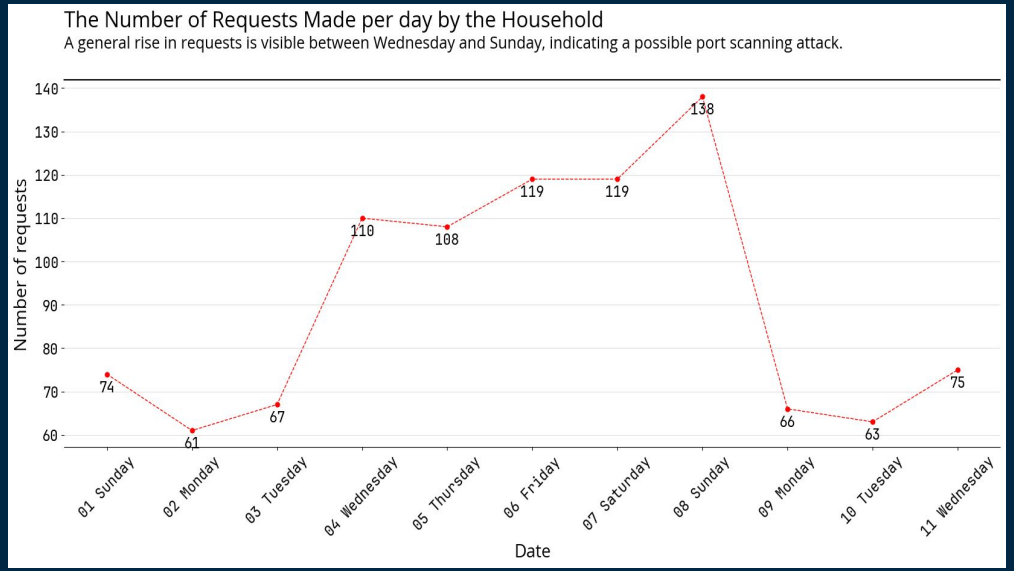
- "Ports Appearing in Requests - Port Scanning" indicates that the graph represents the occurrence of different ports in requests, potentially associated with port scanning activity.
- "A general rise in ports commonly vulnerable to attacks is visible." This statement suggests that there's a noticeable increase in requests targeting ports that are often vulnerable to cyber attacks.
- Ports 21 (FTP), 53 (DNS), 80 (HTTP), and 443 (HTTPS) have relatively higher bars, indicating a higher volume of requests targeting these ports.
- Ports 67 (DHCP), 25 (SMTP), 143 (IMAP), and 110 (POP3) have relatively lower bars, suggesting fewer requests aimed at these ports.
-



- "The Number of Requests Made per day by the Household" indicates that the graph represents the frequency of requests made by the household on each day.
- "A normal and healthy usage of the internet is seen with the occasional spike here and there." This statement suggests that overall, the household's internet usage is within typical bounds, with sporadic spikes in activity.
- The data points vary in their vertical position, indicating fluctuations in the number of requests from day to day.
- There are noticeable spikes in activity on certain days, such as on the 3rd, 7th, and 11th, where the number of requests notably increases compared to the surrounding days.
- Overall, the line graph shows a fluctuating pattern, with some peaks and troughs, suggesting variability in the household's internet usage.



- "The Number of Requests Made per day by the Household" indicates that the graph represents the frequency of requests made by the household on each day.
- "A general rise in requests is visible between Wednesday and Sunday, indicating a possible port scanning attack." This statement suggests that there's a notable increase in requests during the midweek to the weekend period, which could potentially signify a port scanning attack, a type of cyberattack that scans a range of network ports on a target system to identify vulnerabilities.



Challenges and Limitations

- Scalability issues in large networks
- Encryption challenges and solutions
- False positives and false negatives in detection

Citations

1. Adams, Niall M., and Nicholas A. Heard, eds. Data analysis for network cyber-security. World Scientific, 2014.
2. Xin, Yang, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixia Hou, and Chunhua Wang. "Machine learning and deep learning methods for cybersecurity." Ieee access 6 (2018): 35365-35381.
3. Xie P, Li JH, Ou X, Liu P, Levy R. Using Bayesian networks for cyber security analysis. In2010 IEEE/IFIP International Conference on Dependable Systems & Networks (DSN) 2010 Jun 28 (pp. 211-220). IEEE.

THANK YOU !!!